Genetic Epidemiology of *Plasmodium falciparum* Asymptomatic Infections and Antimalarial Drug-resistance Markers in Kilifi, Kenya.

Thesis

Genetic epidemiology of *Plasmodium falciparum* asymptomatic infections and antimalarial drug-resistance markers in Kilifi, Kenya.

Kevin Kariuki Wamae (B.Sc., M.Sc.)

A thesis submitted for the degree of Doctor of Philosophy.

The Open University, UK

Affiliated Research Centre

KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya.

KEMRI | Wellcome Trust

Collaborating Establishment

Infectious Disease Epidemiology and Ecology Lab (IDEEL)

(University of North Carolina at Chapel Hill School of Medicine)

UNC SCHOOL OF MEDICINE

September, 2019

Candidate's Contribution

1. Objective 1 (chapter 2) - Statistical analysis (Studies designed and conducted by collaborators)
2. Objective 2 (chapter 3) - DNA extraction, PCR amplification and Sanger sequencing of three of the twelve *Plasmodium falciparum* antimalarial drug resistance markers (*k13*, *ap2mu* and *falcipain-2a*). The remaining drug resistance markers (*crt*, *mdr1*, *dhps*, *nfs*, *ubp-1*, as well as four artemisinin resistance predisposing mutations; *arps10* codon V127M*, crt* codon I356T*, fd* codon D193Y and *mdr2* codon T484I) were amplified and sequenced by members of Dr. Lynette Isabella Ochola-Oyier's group.
3. Objective 3 (chapter 4) - DNA extraction, PCR amplification and deep-sequencing of *P. falciparum ama1* gene.
4. The analysis, interpretation and presentation of all work contained here is the candidate's work only.

Acknowledgement

I would like to thank the God, in whom I have found the strength, will, health and guidance to make this project a success.

To my family - your love, prayers and emotional support cannot be put into words.

To my supervisors, Dr. Lynette Isabella Oyier, Prof. Philip Bejon and Dr. Colin Sutherland, thank you for offering me the resources that enabled the undertaking of this project as well as your overwhelming guidance and critic.

To my funders, The Initiative to Develop African Research Leaders (IDeAL) led by Dr. Sam Kinyanjui, thank you for offering me a Ph.D. fellowship and unrelenting support during the entirety of this project.

To all members of Dr. Lynette's group including Victor Osoti, Leonard Marigi, Dorcas Okanda, Dr. Mercy Akinyi, Kelvin Muteru and Peninnah Makusa, thank you for your assistance with retrieving study samples, running lab experiments and for the insightful group discussions.

Many thanks to Juliana Wambua, Jennifer Musyoki, Barnes Kitsao, George Nyangweso and Zaydah De Laurent for your assistance in acquiring the malaria monitoring data, study samples and help with conducting lab experiments.

I am grateful to Dr. Jonathan Juliano (University of North Carolina, U.S.A) and Dr. Jefferey Bailey (Brown University, U.S.A) for allowing me to visit their labs and for their help with conducting some of the resource insensitive lab experiments. Your support played a huge role in the completion of this project. Also, my sincere thanks to Dr. Jonathan's lab members for their unrelenting support during my stay in the U.S.A.

To my third-party monitor, Dr. Charles Agoti and my mentor, Prof. Olubayi Olubayi, you offered encouraging words and guidance that enabled me to navigate some of the challenges I encountered.

Finally, many thanks to my colleagues at KEMRI-Wellcome Trust for your relentless critic and support as well as the KEMRI-Wellcome Trust as a whole, you are an addition to my ever-growing family.

Dedication

"To My Family" - A pillar to lean on in times of need. Your love, pushes me and brings out the best in me.

Abstract

Asymptomatic infections and anti-malarial drug resistance are impediments to malaria elimination. Markedly, asymptomatic infections often go undetected and untreated and this creates a parasite reservoir that fuels malaria transmission as well as being a risk factor for febrile malaria. On the other hand, drug-resistance renders antimalarial drugs ineffective and has been associated with increased morbidity and mortality in the past.

Using longitudinal malaria monitoring data and *Plasmodium falciparum* positive samples from Kilifi, Kenya, *(i)* an evaluation was conducted to assess the impact of age and malaria transmission intensity on the risk of developing febrile malaria in individuals harbouring asymptomatic infections, *(ii)* amplicon deep-sequencing was used to evaluate *P. falciparum* genetic diversity in asymptomatic and febrile infections and *(iii*) the diversity of twelve drug-resistance markers (*crt, mdr1, dhps, nfs, k13, ap2mu, falcipain-2a, ubp-1, as well as four artemisinin resistance predisposing mutations; arps10 codon V127M, crt codon I356T, fd codon D193Y and mdr2 codon T484I*) in Kilifi was evaluated using Sanger sequencing, including a neutral marker (*serine-tRNA ligase*) that is not under drug-pressure.

Analyses of the data revealed that in the moderate and high transmission intensity settings, asymptomatic infections were associated with a reduced risk of febrile malaria in older children (>3 years), while in the lower transmission setting, asymptomatic infections were associated with an increased risk of febrile malaria in children of all ages. Amplicon deep-sequencing revealed that *P. falciparum* genetic diversity in asymptomatic and febrile infections differ significantly, similar to previous reports. Also, a majority of the febrile cases (86%) were due to the introduction of *P. falciparum* clones that were not detected in the preceding asymptomatic episode. Regarding the analysis of antimalarial drug resistance markers, none of *kelch 13 (k13)* validated markers of artemisinin resistance were detected in the population, nonetheless, a single *k13 allele*, K189T, was maintained at a stable high frequency (>10%) over time. There was a distinct shift from *chloroquine resistant transporter (crt)*-76, *multi-drug resistant gene 1* (*mdr1)*-86 and *mdr1*-1246 chloroquine (CQ) resistance alleles to a 99% prevalence of CQ sensitive alleles in the population, following the withdrawal of CQ from routine use. In contrast, the *dihydropteroate synthetase (dhps)* double mutant

(437G and 540E) associated with sulfadoxine-pyrimethamine (SP) resistance was maintained at a high frequency (>75%), after a change from SP to artemisinin combination therapies (ACTs). The novel *cysteine desulfurase (nfs)* K65 allele, implicated in resistance to lumefantrine in a West African study, showed a gradual significant decline in allele frequency pre- and post-ACT introduction (from 38% to 20%), suggesting evidence of directional selection in Kenya, potentially not due to lumefantrine. The frequency of AP-2 complex subunit mu (*ap2-mu*) S160N allele, a mutation that has been associated with directional selection after artemisinin combination therapy, was stable over time indicating that it is not under drug-pressure. On the other hand, the ubiquitin carboxyl-terminal hydrolase 1 (*ubp-1)* mutation at codon E1528D, also associated with directional selection after artemisinin combination therapy, was not detected. The S69Stop mutation in *falcipain-2a* that has been associated with artemisinin resistance, *in vitro*, was not detected and none of the artemisinin resistance predisposing mutations were identified.

Asymptomatic infections were found to be modified by transmission intensity and age, altering the risk of developing febrile episodes and this suggested that host immunity plays a prominent role in mediating this process. The differences in *P. falciparum* genetic diversity between asymptomatic and febrile malaria infections can be attributed to the broader spectrum of immunological memory in the form of antibodies that has been found to be higher in asymptomatic infections compared to febrile malaria infections. While evidence of apparent protection from developing febrile episodes was observed in children with asymptomatic infections, amplicon deep-sequencing revealed that this protection was offset when a new infection occurred. Lastly, due to lack of the validated molecular markers of artemisinin resistance, there appears to be no problem of resistance in the population, however, continued surveillance remains a requirement. To conclude, *P. falciparum* genetic epidemiology revealed the ability to characterise *P. falciparum* complexity of infection (COI), a proxy that can be used to characterise malaria transmission intensity. Additionally, frequent sampling of *P. falciparum* positive samples from a hospital setting coupled with the molecular genotyping of drug resistance markers revealed the utility of *P. falciparum* genetic epidemiology in surveillance of antimalarial drug resistance.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACT - artemisinin-based combination therapies

AL - artemether-lumefantrine

ama1 - apical membrane antigen 1

An. gambiae - *Anopheles gambiae*

ap2-mu - AP-2 complex subunit mu gene

arps10 - apicoplast ribosomal protein S10 precursor

AS-AQ - artesunate-amodiaquine

AS-MQ - artesunate-mefloquine

AS-SP - artesunate + sulfadoxine-pyrimethamine

COI - complexity of infection

CQ - chloroquine

CQR - chloroquine resistance

CQS - chloroquine sensitive

crt - chloroquine resistance transporter

DHA-PPQ - dihydroartemisinin-piperaquine

dhfr - dihydrofolate reductase

dhps - dihydropteroate synthase

EIR - entomological inoculation rate

fd - ferredoxin

G6PD - glucose-6-phosphate dehydrogenase

glurp - glutamine-rich protein

IRS - indoor residual spraying

ITN - insecticide-treated mosquito net

k13 - kelch 13 protein

kdr - knockdown resistance

MDA - mass drug administration

mdr - multidrug resistance

MQ - mefloquine

msp - merozoite surface protein

NAI - naturally acquired immunity

nfs - cysteine desulfurase

P. falciparum - *Plasmodium falciparum*

PCR - polymerase chain reaction

PfEMP1 - *P. falciparum* erythrocyte membrane protein 1

RDT - rapid diagnostic tests

RSA - ring-stage survival assays

RBC - red blood cell

SEA - South-East Asia

SNP - single nucleotide polymorphism

SP - sulfadoxine and pyrimethamine

SSA - sub-Saharan Africa

WHO - World Health Organisation

ubp-1 - ubiquitin carboxyl-terminal hydrolase 1

*μ*l - microlitre

1

Chapter 1 : General Introduction

1.1 Background

Malaria in humans is caused by six species of unicellular protozoan parasite of the genus *Plasmodium (*the species include: *falciparum*, *vivax*, *malariae*, *ovale curtisi, ovale wallikeri* and *knowlesi)* and it is considered one of the most important infectious diseases still affecting mankind today. Moreover, while nearly half of the world's population is at risk of malaria, sub-Saharan Africa (SSA) continues to bear the greatest burden of *Plasmodium falciparum* malaria, the species associated with the most serious morbidity and mortality worldwide (WHO 2018a). For this reason, efforts to reduce or eliminate malaria burden in SSA remain a priority. Between 2000 and 2015, the implementation of infection control measures such as insecticide-treated mosquito nets (ITNs) and artemisinin-based combination therapies (ACTs) in SSA resulted in the aversion of more than 663 million clinical cases and over 6 million deaths (Bhatt *et al.* 2015). However, while significant progress has been made, equally significant challenges persist. Notably, In 2017 there were an estimated 219 million malaria cases and approximately 435,000 deaths globally and SSA accounted for most of the global malaria cases (92%) and deaths (93%) (WHO 2018a). Therefore, this calls for more research to be done to inform malaria elimination efforts.

1.2 The life cycle of *P. falciparum*

P. falciparum, the focus of this work, is the most prevalent *Plasmodium* species in SSA and is responsible for the greatest burden of malaria (WHO 2018a). The *P. falciparum* life cycle commences when an infected female *Anopheles* mosquito injects *Plasmodium* sporozoites into the host during a blood meal (**Figure 1.1**). Some sporozoites then enter the host's blood circulation or lymphatic system and traffic to the liver. There, the sporozoites cross the sinusoidal layer, primarily through Kupffer cells to access hepatocytes - liver cells (Pradel and Frevert 2001). Sporozoites migrate through several hepatocytes before invading a final hepatocyte, in which a parasitophorous vacuole is formed (Mota *et al.* 2001). Here, the sporozoites develop by a process of schizogony into merozoite forms. The parasites then induce death and detachment of their host hepatocytes, followed by the budding of parasite-filled vesicles (merosomes) into the sinusoid lumen (Sturm *et al.* 2006).

Approximately 30,000 merozoites rupture from an infected liver cell to invade red blood cells (RBCs). Some *Plasmodium spp.* such as *P. vivax* also develop into a small number of latent stage forms called "hypnozoites". These latent-stages can revert to active infective parasites that cause clinical relapses of malaria within weeks or months of the original infection (Cogswell 1992). Upon the invasion of RBCs, merozoites first develop into 'ring' trophozoites, then mature into trophozoites and finally into a schizont containing up to 32 new merozoites. These schizont-infected RBCs synchronously rupture to release merozoites, which can invade new RBCs, resulting in an exponential increase in parasite biomass. This multiplication and invasion of RBCs lead to illness and complications of malaria.

The *Plasmodium* life cycle continues when some merozoites develop into the sexual parasite stages, male and female gametocytes, the form that is taken up by the mosquito host (Silvestrini, Alano and Williams 2000; Smith *et al.* 2000). The mechanism underlying commitment to sexual differentiation in *P. falciparum* was only recently shown to be an epigenetically regulated process. Two recent studies showed that the AP2 domain transcription factor (AP2-G) is a key regulator of gametocytogenesis, as in the absence of AP2-G parasites do not produce gametocytes, both in *P. falciparum* (Kafsack *et al.* 2014) and its murine relative, *P. berghei* (Sinha *et al.* 2014). Following gametocytogenesis, male and female gametocytes emerge from within their infected erythrocytes and undergo fertilisation, zygote formation and development into the motile ookinete which infect the mosquito's midgut (Angrisano *et al.* 2012). From here, they develop into oocysts that burst to release sporozoites which invade the mosquito's salivary glands awaiting another blood meal for the cycle to continue.

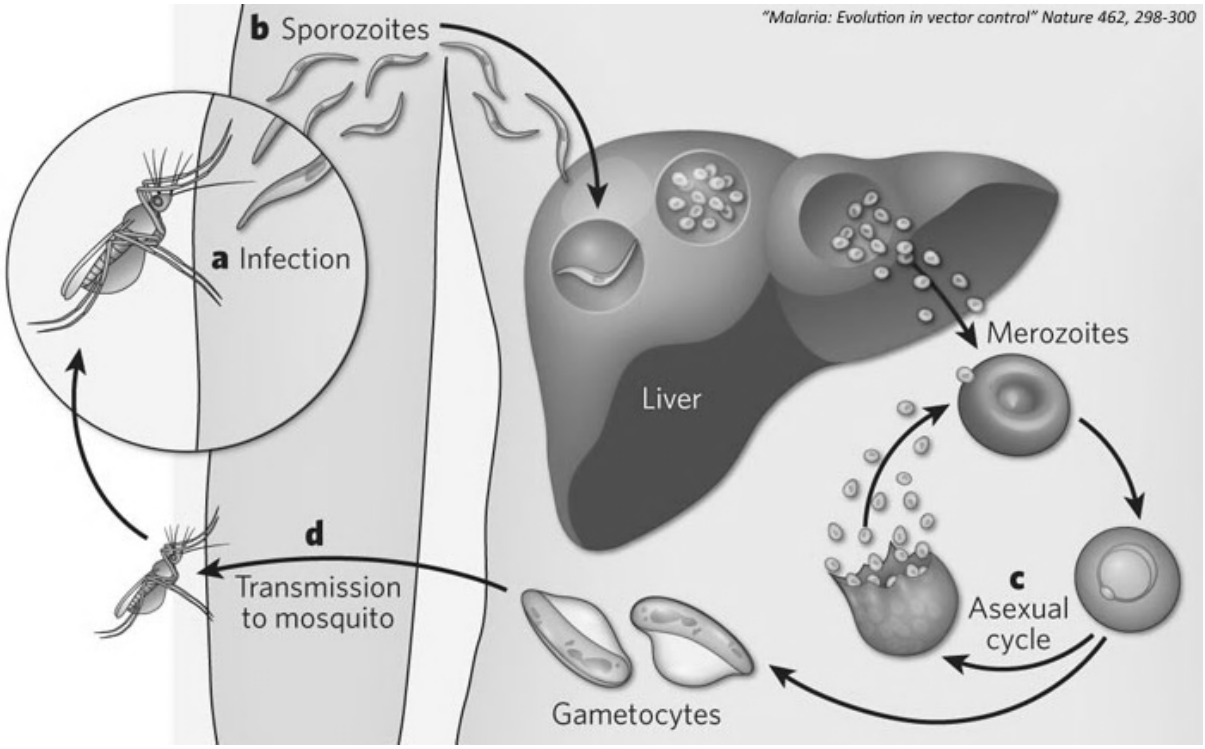


Figure 1.1. The *P. falciparum* life cycle. The figure shows the cycle in the human host which commences with sporozoites invading the liver, the sporozoites then mature and rapture to release merozoites. These merozoites invade red blood cells to begin the asexual cycle, accompanied by clinical manifestations of malaria. This is followed by the production of male and female gametocytes which are transmitted back to the mosquito (during a blood meal) where they fuse to form oocysts that divide to create sporozoites. The sporozoites migrate to the mosquito's salivary glands from where the cycle of infection starts again. *Adapted from Michalakis and Renaud (2009)*.

1.3 Laboratory diagnosis

The WHO recommends the confirmation of the presence of malaria parasites in suspected malaria cases before antimalarial treatment. Consequently, microscopy of stained blood films remains the gold standard for diagnosis of malaria (WHO 2015). However, rapid diagnostic tests (RDTs) are now widely used as the first-line diagnostic tool and given the predominance of *P. falciparum* in Africa, a *P. falciparum*-only RDT is often used, which is based on the histidine-rich protein 2 (HRP2) specific to *P. falciparum*. Nonetheless, some RDTs include an additional *Plasmodium* lactate dehydrogenase (p-LDH) band for the detection of non-*falciparum* species (Chiodini 2014). Recent reports of *P. falciparum* HRP2 and HRP3 gene

deletions provide new obstacles in the path of malaria elimination and warrant regular surveillance for such mutations (Verma, Bharti and Das 2018).

Microscopy, the gold standard, typically detects 50 parasites/μl in European reference laboratories (Bejon *et al.* 2006). However, this may vary in the field where detection may range between 50 and 100 parasites/μl because of lower quality reagents and less stringent protocols for slide preparation (Wongsrichanalai *et al.* 2007). RDTs show a sensitivity similar to that of good microscopy, approximately 100 parasites/μl, but even these may fail to detect lower density infections (Wongsrichanalai *et al.* 2007). Therefore, nucleic acid amplification-based tests offer greater sensitivity and can reach limits of 0.02 parasites/μl (Mahajan *et al.* 2012). That said, microscopy remains the most affordable diagnostic method for diagnosis, especially in the resource-limited settings of Africa.

1.4 Manifestations of malaria

Manifestation of *P. falciparum* infection ranges from asymptomatic parasitaemia, uncomplicated malaria, through to severe malaria. Asymptomatic malaria infections can be defined as infections that lack typical clinical malaria symptoms but whose parasitemia is detectable by microscopy, rapid diagnostic tests (RDTs) or molecular methods (Bousema *et al.* 2014). Uncomplicated malaria infections typically present with fever and without signs of severity or evidence (clinical or laboratory) of vital organ dysfunction and since the signs and symptoms of uncomplicated malaria are nonspecific, it is clinically diagnosed based on fever or a history of fever (WHO 2015). This, however, presents a challenge in distinguishing uncomplicated malaria from Dengue fever and Chikungunya (viral infections) since clinical presentations tend to be shared, these infections are co-endemic both in SSA and South-East (SE) Asia and fever is the most common symptom (Salam *et al.* 2018). Severe *falciparum* malaria presentations are varied and include the following:

- cerebral malaria - unrousable coma with peripheral *P. falciparum* parasitaemia after exclusion of other causes of encephalopathy.
- severe anaemia - normocytic anaemia with haemoglobin <5g/dl (or haematocrit <15%)

- respiratory distress - pulmonary oedema or acute respiratory distress syndrome
- renal failure - urine output of less than 400 ml in 24 h (or <12 ml/kg in children) and a serum creatinine >265 mmol/l (>3.0 mg/dl).
- Hypoglycaemia - whole blood glucose <2.2 mmol/l (40 mg/dl).
- circulatory collapse (shock) - systolic blood pressure <70 mmHg or core-skin temperature difference >10°C.
- coagulation failure - spontaneous bleeding and/or laboratory evidence of disseminated intravascular coagulation (WHO 2000).

The malaria disease spectrum is quite complex and is influenced by factors related to the host, parasite, vector and the environment. In order to effectively implement anti-malaria interventions, a deeper understanding of the epidemiology of malaria is crucial.

1.5 Epidemiology of *falciparum* malaria

1.5.1 The malaria burden

According to the World Malaria Report of 2018 (WHO 2018a), 20 million fewer malaria cases were reported in 2017 than in 2010, but there was no significant progress in reducing global malaria cases for the period 2015-2017. Fifteen countries in SSA and India carried almost 80% of the global malaria burden and the 10 highest-burden countries in SSA reported increases in cases of malaria in 2017 compared with 2016. In contrast, India reported 3 million fewer cases in the same period, a 24% decrease compared with 2016. While SE Asia continued to witness a decline in incidence rates (from 17 cases per 1000 population at risk in 2010 to 7 in 2017, a 59% decrease), the malaria incidence rate remained at 219 cases per 1000 population at risk in SSA for the second year in a row. *P. falciparum* remained the predominant parasite in SSA and accounted for 99% and 62% of estimated malaria cases in 2017 in SSA and SE Asia, respectively. In comparison, *P. vivax* was the predominant parasite in the South Americas, representing 74% of malaria cases. Approximately 80% of global malaria deaths in 2017 were concentrated in 17 countries in the SSA and India. Conversely, while SSA experienced the highest number of malaria deaths in 2017, it experienced 151,000 fewer malaria deaths compared with 2010 (WHO 2018a).

1.5.2 Malaria endemicity

Malaria endemicity can be used to indicate disease prevalence and may help to design, implement and monitor malaria control and prevention activities. Various methods exist for classifying malaria endemicity (Autino *et al.* 2012), including *(i)* the proportion of individuals in a population with a substantial enlargement of spleen (spleen rate), *(ii)* the proportion of individuals in a population with a laboratory-confirmed parasite infection (parasite rate), *(iii)* the number of infective bites per person (entomological inoculation rate) and *(iv)* the number of microscopically confirmed malaria cases detected during one year per unit population (annual parasite incidence). Of these, EIR best quantifies endemicity as it is more directly related to morbidity and mortality than the rest, however, measuring EIR requires intensive entomological studies over the whole annual period of malaria transmission (Beier, Killeen and Githure 1999). Though variable, SSA has the highest annual EIRs that range from 0.6 to 814 (Kelly-Hope and McKenzie 2009) compared to SE Asia which has low EIRs such as 0.3 in India, 2.2 in Vietnam, 3 in Thailand and 0.13 in Myanmar (Chattopadhyay *et al.* 2004; Van Bortel *et al.* 2010; Kwansomboon *et al.* 2017; Chaumeau *et al.* 2019).

How does the burden of malaria vary with malaria transmission intensity? In a study by Carneiro et al. (2010) that included 86 studies from SSA, episodes of uncomplicated malaria were fairly evenly distributed across all ages and this burden shifted towards younger age groups as transmission intensity increased. Hospital admissions due to *falciparum* malaria were found to be concentrated in younger children, as did severe cases of malaria, and these were more likely with increasing transmission intensity. Moreover, malaria-diagnosed mortality was more focused in younger children than admissions with malaria in all settings. All three outcomes, uncomplicated malaria, hospital admissions with *falciparum* malaria and malaria-diagnosed mortality, were biased towards younger ages.

In a similar study, Roca-Feltrer et al. (2010) evaluated the relationship between age patterns and severe malaria syndromes (cerebral malaria - CM, severe malarial anaemia - SMA and respiratory distress - RD) from 67 studies in SSA. They observed that increasing transmission intensity shifted the burden of CM towards younger age groups however, this was not apparent for SMA or RD. for this reason, the authors concluded that there are many causes

of SMA, including HIV and malnutrition, which may mask the relationship between transmission intensity and SMA. On the other hand, admissions of RD frequently occur with SMA and CM and like SMA, this may mask the relationship between transmission intensity and SMA.

This shift in the malaria burden as a consequence of malaria transmission intensity is a result of the rapid acquisition of immune responses that limit the life-threatening effects of malaria due to increasing exposure to *P. falciparum* (Doolan, Dobaño and Baird 2009). Nevertheless, while declines in malaria transmission intensity have been associated with a shift in malaria morbidity towards older children due to reduced exposure, strategies for averting malaria mortality and severe morbidity in young children are still appropriate.

1.5.3 The distribution of *Anopheles* species

The *Anopheles gambiae complex* is the most effective and efficient of *P. falciparum* vectors and predominates if SSA, leading to some of the highest EIRs and the highest malaria prevalence worldwide (Guerra *et al.* 2008). Four principal species of *An. gambiae complex* exist, including *An. gambiae, An. arabiensis, An. merus* and *An. melas*. However, three additional anthropophilic vectors exist in SSA including *An. funestus, An. moucheti* and *An. nili* (Sinka *et al.* 2010). Conversely, *An. Dirus* and *An. darlingi* are considered the most important malaria vectors in SE Asia and South America, respectively (Obsomer, Defourny and Coosemans 2007; Sinka *et al.* 2010).

Vector distribution is influenced by environmental factors such as rainfall patterns, temperature, humidity, presence of vegetation and surface water and hence impact on the malaria transmission cycle. Human activities such as agriculture, irrigation, deforestation, urbanization and population movements are also connected to transmission levels and malaria epidemiology (De Silva and Marshall 2012), further revealing how complex the malaria transmission cycle is and the need for a multi-faceted approach to tackling malaria.

1.5.4 The distribution of *Plasmodium* species

P. falciparum is prevalent in virtually all malaria-endemic countries. In SSA, *P. falciparum* is the most dominant species with relative frequencies as high as 98% of malaria infections

(Culleton *et al.* 2008). *Vivax* malaria is less prevalent in SSA because many individuals are Duffy antigen-negative, however, there have been reports of *vivax* infections in Duffy-negative individuals in SSA (Zimmerman 2017). Thus, whether *vivax* malaria was undetected in SSA in the past or the parasite has adapted to infect Duffy negative individuals, it remains to be seen how significant *vivax* malaria will be in SSA. In contrast, numbers of malaria incidences in SE Asia and Oceania are generally lower and are caused by both *P. vivax* and *P. falciparum*, whereas in South America, *vivax* malaria cases exceed *falciparum* by more than two-fold (WHO 2018a).

The *P. ovale* species is relatively less prevalent compared to other malaria parasites, occurring in <5% of cases associated with the parasite, however, higher prevalence has been observed for example in Cameroon where one study reported a prevalence greater than 10% (Collins and Jeffery 2005). *P. malariae* is spread in SSA, SE Asia and South America and its distribution overlaps with that of *P. falciparum* (Collins and Jeffery 2007). Still, the distribution of *P. malariae* is relatively sparse, variable and is endemic to West-Africa (Roucher *et al.* 2014), South America (Scopel *et al.* 2004), SE Asia (Zhou *et al.* 1998; Mohapatra *et al.* 2008) and the western Pacific region (Mueller *et al.* 2005). *P. knowlesi* is a parasite of the long-tailed and pig-tailed macaques (Vythilingam *et al.* 2008), however, it can also infect humans and infections are localized in SE Asia because the long-tailed and pig-tailed macaques inhabit the forests of SE Asia (Mason Dentinger 2016).

1.6 Malaria case management

For uncomplicated *P. falciparum* malaria, the WHO (WHO 2015) recommends treatment of children and adults (except pregnant women in their first trimester) with one of the following ACTs: artemether-lumefantrine (AL), artesunate-amodiaquine (AS-AQ), artesunate-mefloquine (AS-MQ), dihydroartemisinin-piperaquine (DHA-PPQ) and artesunate + sulfadoxine-pyrimethamine (AS-SP). In low-transmission areas, a single low dose of 0.25 mg/kg primaquine with an ACT is recommended for patients with *P. falciparum* malaria (except pregnant women, infants aged <6 months and women breastfeeding infants aged <6 months) to reduce transmission without the need for glucose-6-phosphate dehydrogenase (G6PD) testing. Even though the use of primaquine in G6PD deficient patients has been

associated with haemolysis, a single low dose is associated with acceptable haemolytic risk (Pamba *et al.* 2012).

Malaria in the first trimester of pregnancy is treated with quinine and clindamycin. However, in areas where clindamycin is unavailable, quinine monotherapy is prescribed. For second and third trimesters, AL, AS-AQ, AS-MQ and DHA-PPQ are recommended, though AL is associated with less adverse effects and DHA-PPQ is associated with the greatest post-treatment prophylaxis (WHO 2015).

Patients presenting with severe malaria are treated with parenteral artesunate (WHO 2014a). Additional measures include urgent correction through blood transfusion for severe anaemia (haemoglobin <6g/dl) (Maitland *et al.* 2019). Patients with cerebral malaria should have blood cultures taken, a lumbar puncture performed and administered with broad-spectrum antibiotics, pending negative culture results and clinical improvement (WHO 2013). For patients presenting with respiratory distress (acidosis), reversible cause of acidosis (dehydration and severe anaemia) are corrected by intravenous infusion, an excess of which may precipitate pulmonary oedema. Hypoglycemia should be corrected with 500mg/kg of glucose and monitored even after successful correction, as hypoglycemia may recur (WHO 2013). Circulatory collapse (shock) is corrected with fluids at 3–4ml/kg per hour, blood is taken for culture and the patient is put on appropriate broad-spectrum antibiotics immediately (WHO 2013).

1.7 Immunity against malaria

Repeated exposure to infectious mosquito bites leads to the acquisition of naturally acquired immunity (NAI), however, this protection is only limited to the disease and does not clear the blood-stage infection (Doolan et al. 2009). Nevertheless, sterile protection can be achieved against a homologous challenge by the inoculation of sporozoites (Roestenberg *et al.* 2009). NAI protects one from severe disease and death and is acquired within the first five years of life in areas of high transmission such as in SSA, whilst immunity to mild disease is acquired by late adolescence (Marsh and Kinyanjui 2006). Consequently, children tend to be protected from severe life-threatening malaria and as they grow older there is an establishment of

chronic, asymptomatic infection, to which even individuals growing up in malaria-endemic areas remain susceptible to, for life (Doolan et al. 2009). Independent of prior exposure, adults acquire NAI to malaria faster than children, as was seen in a study conducted in Indonesia that compared immune responses between native residents of a *P. falciparum*-endemic region and trans-migrants following their re-settlement from non-*P. falciparum*-endemic islands (Baird *et al.* 1993). Hence, these findings suggest that NAI to malaria is not only a function of exposure but at least partly attributable to inherent age-specific host factors, such as a mature or fully developed immune system.

In *P. falciparum*, NAI has been associated with the sequential acquisition of antibodies and *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1), a protein expressed on the surface of parasitized erythrocytes, has been implicated as the key target (Dodoo *et al.* 2001). This sequential acquisition of antibodies is the result of cumulative exposure to multiple parasite infections over time, yielding a diverse repertoire of strain-specific immune responses. In fact, *P. falciparum* antigens, the targets of NAI, are highly diverse. In a study conducted in Mali to assess the genetic diversity of apical membrane antigen 1 (*ama1*), Takala et al. (2009) identified as many as 214 distinct *ama1* variants in *P. falciparum* infections experienced by 100 individuals over 3 years. This poses a challenge in the design of a malaria vaccine because of the immense amount of antigenic diversity to be dealt with, hence there is a need to identify immunodominant targets that are likely to be more conserved.

In high transmission settings such as in SSA, it may take 10 to 15 years of roughly five infections per year to develop NAI to malaria (Doolan et al. 2009). However, this immunity is lost gradually after an individual leaves an endemic area or in a population with falling transmission (Ghani *et al.* 2009). Additionally, while reducing malaria transmission is a major component of malaria elimination efforts, there appears to be a detrimental impact. Recent studies have associated declines in malaria transmission with higher rates and severity of clinical malaria (O'Meara *et al.* 2008; Griffin, Ferguson and Ghani 2014; Njuguna *et al.* 2019), something that may be attributed to declining NAI in populations. Considering this, it is important to prevent a re-introduction of malaria in an area that has managed to interrupt malaria transmission to prevent catastrophic malaria epidemics.

1.8 Malaria prevention and control

According to the latest high-resolution *falciparum*-malaria maps for the period 2000-2017 (Weiss *et al.* 2019), the world has witnessed a rapid decline in malaria burden between 2005 and 2017 with clinical episodes declining by 27% and mortality declining by 42%. Notwithstanding a growing population in malaria-endemic regions, clinical cases declined between 2005 and 2017, from 232 million to 193 million. On the other hand, malaria-related deaths declined from 925,000 to 618,000, demonstrating the progress made in combating malaria. Still, this progress appears to have stalled between 2015 and 2017 as there was an increase in clinical cases from 214 million in 2015 to 219 million in 2017, globally. Novel interventions are therefore needed to reverse the curve and accelerate the decline in the malaria burden.

Several factors can be attributed to the recent decline in malaria burden including, but not limited to, vector control, mass drug administration (MDA), as well as early diagnosis and treatment. The two main vector control strategies, ITNs and indoor residual spraying (IRS), are used to target mosquitoes indoors, but do not prevent biting for vectors that prefer to feed outdoors (Kenea *et al.* 2016). ITNs and IRS act by breaking the chain of transmission between the mosquito vector and the human host and both have had remarkable success in reducing both the malaria parasite prevalence and incidence across a range of transmission settings. IRS has been associated with significant disruption in malaria transmission, is known to eliminate malaria vectors and has enabled the reduction of malaria incidences (Pluess *et al.* 2010). With regards to ITNs, results from a systematic review using national cross-sectional survey datasets in SSA showed that ITN possession was associated with a 20% reduction in *P. falciparum* prevalence (from seven surveys in seven countries) and a 23% reduction in all-cause child mortality (from 29 surveys in 22 countries) (Lim *et al.* 2011). Similar findings were reported in a more recent Cochrane review that included data from 23 trials enrolling more than 275,793 adults and children (1987 and 2001) (Pryce, Richardson and Lengeler 2018). ITN use was associated with a 17% reduction in all-cause child mortality and also reduced the incidence of uncomplicated clinical episodes by half. ITNs were also associated with a 17% reduction in parasite prevalence as well as a 44% reduction in severe malaria episodes.

Nevertheless, ITN coverage remains insufficient and efforts need to be stepped up to ensure universal coverage (WHO 2018a). Additionally, insecticide resistance to the four classes of insecticides (pyrethroids, organochlorines, organophosphates and carbamates) presents a major obstacle to vector control (Coleman *et al.* 2017).

MDA involves the administration of a full dose of antimalarial treatment to an entire population in a given area, regardless of the symptoms or presence of infection - except those in whom the medicine is contraindicated. Also, to reduce malaria transmission, MDA aims to provide therapeutic concentrations of antimalarial drugs to as large a proportion of the population as possible to cure asymptomatic infections and to prevent re-infection during the period of post-treatment prophylaxis (WHO Global Malaria Programme 2015). MDA has been used for malaria control since the 1930s and was backed by WHO for malaria elimination and eradication in the 1950s (Poirot *et al.* 2013). Moreover, MDA has been shown to be successful in combination with other malaria control measures, such as vector control. For example, MDA with SP combined with IRS achieved high levels of malaria control in Northern Nigeria in 1969 (Molineaux and Gramiccia 1980), although the effect was short-lived. More recently, a review of published literature on MDA by Poirot et al. (2013), included areas of different endemicity, various antimalarials and dosages, different timings and the number of MDA rounds as well as the simultaneous implementation of vector-control measures. The review concluded that MDA appears to reduce malaria parasitaemia rapidly plus several clinical outcomes. However, more studies are required to assess the impact after 6 months, assess the barriers for community uptake and explore the potential impact on the development of drug resistance. To date, MDA has not been adopted as official policy in any SSA country.

The rationale for early diagnosis and treatment (EDT) is to offer a means of reducing transmission of *P. falciparum* from febrile individuals. Thus, treating *P. falciparum*-infected patients after the onset of fever is expected to prevent further transmission of the parasite (Landier *et al.* 2016). Equally, a delay in treatment means that a patient is more likely to remain infectious even after being treated because gametocytes, the transmissible sexual forms of *P. falciparum*, may persist for several weeks even after clearance of asexual parasites (Bousema and Drakeley 2011). Nonetheless, the inclusion of a single dose of 0.25mg/kg

primaquine to ACT regimens has been shown to target gametocytes and reduces *P. falciparum* transmission (Graves *et al.* 2018). In a review by Landier et al. (2016), a community-based EDT for malaria is made up of three main components including trained personnel, quality RDT and effective anti-malarial drugs. The continuous availability of these tools is expected to provide reliable diagnosis and effective treatment for any clinical malaria case from the community within 24–48 hours of fever onset. While a functioning EDT network embodies an essential component of any malaria elimination strategy, it may be unavailable in resource limited settings, such as those in SSA, and efforts should be geared towards establishing and sustaining them.

1.9 Malaria elimination

While the idea of eliminating malaria presents methodical challenges, it comes with vast economic, equity and health benefits. Certification of malaria elimination by the WHO is the official recognition of a country being free of indigenous malaria cases, based on an independent evaluation verifying interruption of transmission and the country's ability to prevent re-establishment of transmission for all human malaria parasites. Additionally, a country must report zero indigenous cases of malaria for 3 consecutive years (WHO 2018a). To achieve global malaria elimination, there is a need for advancements that will transform the malaria community's ability to tackle the disease.

Vector control, active surveillance and efficient case management have been instrumental in setting several countries on the path to eliminating malaria. Between the years 2000 and 2017, 19 countries (three in Africa including Algeria, Egypt and Morocco) attained zero indigenous cases for 3 years or more and 21 countries (six in Africa including Algeria, Botswana, Cape Verde, Comoros, Eswatini and South Africa) have been identified as having the potential to eliminate malaria by 2020 (WHO 2018a). With the prevalence of malaria being heterogeneous both locally and across geographic regions, a single approach aimed at eliminating the disease will not work across all settings with the same effectiveness. Therefore, the path to malaria elimination as recommended by the WHO (WHO 2017) is organised into four components: firstly, enhancing and optimizing vector control and case management by providing universal access to malaria prevention, diagnosis and treatment in

all areas. Secondly, increasing the sensitivity and specificity of surveillance to tackle malaria cases and respond appropriately. Thirdly, accelerating transmission reduction in which interventions such as MDA are included and lastly, once very low intensity of transmission has been achieved, a country must be capable of finding the few remaining infections and any foci of ongoing transmission, identify them and clear them.

Even with the recent reports of declining *falciparum* malaria transmission globally, concerted efforts are still needed to achieve complete elimination of malaria in SSA and SE Asia. This is because many regions in SSA and SE Asia still lack significant malaria-free areas and transmission rates remain substantial even after several years of continuous intervention programmes. Also, while gains have been made, several challenges persist that pose challenges to malaria elimination, including insecticide resistance, antimalarial drug resistance, chronic asymptomatic infections and extensive parasite genetic diversity and these are discussed below.

1.9.1 Insecticide resistance

Insecticide resistance can be defined as the ability of mosquitoes to survive exposure to a standard dose of insecticide as a result of physiological or behavioural adaptation. Resistance to insecticides (pyrethroids, organochlorines, organophosphates and carbamates) is prevalent in *Anopheles* mosquitoes in all malaria-endemic countries, especially resistance to pyrethroids, the class of insecticides used on all ITNs (WHO 2018a).

The mechanisms by which mosquitoes develop resistance to insecticides can be grouped into four distinct categories (Corbel and N'Guessan 2013) including metabolic resistance, target-site resistance, reduced absorption and behavioural avoidance with the most common mechanism being metabolic resistance. This form of resistance is based on the enzyme systems which enables insects to detoxify naturally occurring insecticides and can result in high levels of insecticide resistance. The second most common resistance mechanism is target-site resistance. Insecticides largely act at a specific site within the insect, usually within the nervous system and the site of action can be modified in resistant strains of insects, hence the insecticide no longer binds effectively. Reduced penetration of insecticides involves

modifications in the insect cuticle or digestive tract linings that prevent or slow the absorption of insecticides. Behavioural changes, that are not based on biochemical mechanisms, enable insects to escape lethal doses of insecticides in response to prolonged exposure.

Regular monitoring for insecticide resistance is essential as it can negatively impact vector control. There are three main ways for detecting insecticide resistance including bioassays, biochemical and molecular tests (WHO 2016). Bioassays involve the exposure of samples of insects from a wild population to a fixed concentration of insecticide on test papers, so that any survivors may be assumed to be resistant. Biochemical tests have now been supplemented with molecular tests that are presently only available for target-site resistance, and resistance to pyrethroids (used in all ITNs) is associated with genetic mutations in the para-type sodium channel gene. These mutations cause reduced neuronal sensitivity and are termed knockdown resistance (*kdr*). Two mutations, L1014F prevalent in W. Africa and L1014S in E. Africa have been reported in pyrethroid-resistant *An. gambiae* (Torre *et al.* 2001; Stump *et al.* 2004; Awolola *et al.* 2005).

It is worrying that pyrethroids are gradually becoming less effective at killing mosquitoes (Ranson and Lissenden 2016) and it is predicted that this drop in effectiveness could lead to increased malaria incidences (Churcher *et al.* 2016). However, while insecticide resistance may negatively impact on the use IRS, it appears that ITNs are still effective in regions where insecticide resistance has been documented (Kleinschmidt *et al.* 2018). Still, the lack of effective insecticides may hamper vector control interventions.

1.9.2 Anti-malarial drug-resistance

Antimalarial drugs have played an important role in the treatment and control of malaria. The major classes of antimalarial drugs include (Farooq and Mahajan 2004) quinoline that was developed from the structural modification of quinine and include 4-aminoquinoline compounds such as chloroquine (CQ) and mefloquine (MQ). The second class of drugs include antifolates such as proguanil, chloroproguanil, pyrimethamine and trimethoprim and sulfa drugs like dapsone, sulfalene, sulfamethoxazole and sulfadoxine. These drugs are used in combinations and the most widely used combination is sulfadoxine and pyrimethamine (SP).

The third class of drugs include artemisinin compounds (artesunate, artemether, arteether), components of the current first-line drugs, ACTs, are synthesised from the plant *Artemisia annua*. In 1993 Malawi became the first African country to replace CQ with SP (Kublin *et al.* 2003) and other countries followed soon after including Kenya in 1999 (Shretta *et al.* 2000). However, following extensive SP resistance, ACTs replaced SP as the first-line antimalarial drug in all malaria endemic countries including Kenya in 2004 (Amin *et al.* 2007).

Extensive use of antimalarials has exerted selective pressure on *P. falciparum* and caused the spread of resistant parasites with increased malaria morbidity and mortality (Trape *et al.* 1998; Korenromp *et al.* 2003). To date, resistance has been reported to all the commonly used antimalarials, in particular, reduced efficacy of former first-line drugs (CQ and SP) and current first-line drugs, ACTs. The identification of *P. falciparum* genetic mutations associated with antimalarial drug-resistance has provided molecular markers for surveillance of resistance, both in real-time and retrospectively, to assess geographic origins and migration patterns of drug-resistant parasites (Cravo, Napolitano and Culleton 2015).

1.9.2.1 Chloroquine (CQ) resistance

Chloroquine resistance (CQR) emerged independently in SE Asia (at the Thai-Cambodian border) and South America (in Colombia) in the late 1950s and then spread to all malaria-endemic regions (Dondorp *et al.* 2010). Although the mechanism of CQ is yet to be understood, it is believed to act by binding to heme molecules in the parasite food vacuole. Heme, a by-product of haemoglobin digestion, is toxic to the parasite hence it is detoxified into hemozoin, which is recognized as 'malaria pigment' by light microscopy. By binding to heme, CQ interferes with the detoxification process, leading to the killing of parasites (Orjih, Ryerse and Fitch 1994). CQ resistance has been associated with reduced CQ accumulation in the food vacuole, hence, CQ sensitive parasites (CQS) accumulate much more CQ in the digestive vacuole than CQR parasites. Consequently, CQ sensitive parasites (CQS) accumulate much more CQ in the digestive vacuole than CQR parasites (Saliba, Folb and Smith 1998). Moreover, point mutations in the genes encoding for the *P. falciparum* chloroquine resistance transporter (*Pfcrt*) and *P. falciparum* multidrug resistance gene (*Pfmdr1*) have been associated with CQ resistance and these are detailed below.

The *P. falciparum* chloroquine resistance transporter (*crt*) is the primary target gene for CQR by extensive mapping of a genetic cross between CQ sensitive (CQS) and resistant clones (Fidock *et al.* 2000; Djimdé *et al.* 2001). The *crt* gene encodes a protein consisting of 424 amino acids that is localized to the parasite's food vacuole and has 10 predicted transmembrane domains. A mutation at codon 76 (lysine to threonine, K76T) is associated with CQR in *P. falciparum* clones from wide geographic areas (Fidock *et al.* 2000) and it is thought that increased efflux of CQ from the food vacuole mediated by *crt* might be associated with the reduction of CQ concentrations in *P. falciparum* resistant isolates (Sanchez *et al.* 2005).

The geographical spread of CQR was described using polymorphic microsatellite markers flanking the *crt* gene (Wootton *et al.* 2002). It was revealed that a single resistant *crt* haplotype, spanning codons 72-76 (CVIET) spread to very high frequencies throughout most of Asia and Africa while the SVMNT haplotype emerged independently and spread in South America. The withdrawal of CQ from widespread use resulted in the resurgence of the CQS haplotype (CVMNK), in areas such as Malawi (Kublin *et al.* 2003) and Kenya (Okombo *et al.* 2014). This has been attributed to a reduction in CQ pressure at a fitness cost to resistant parasites due to enhanced survival of drug-susceptible parasites.

The *P. falciparum* multidrug resistance gene (*pfmdr1*) was identified as the second candidate CQ drug-resistant marker in 1989 (Foote *et al.* 1989; Wilson *et al.* 1989). The gene encodes a homologue of the mammalian multidrug resistance gene, termed P-glycoprotein homolog 1 (Pgh1) (Cowman 1991). Five polymorphisms occurring at codons N86Y, Y184F, S1034C, N1042D and D1246Y have been associated with CQR (Foote *et al.* 1990). However, earlier studies showed that *mdr1* mutants might not directly cause CQR, but may augment CQR since transfection of *mdr1* mutants (at codons 86, 1034, 1042 and 1246) into a susceptible CQS clone did not produce a pronounced change in response to CQ (Haruki *et al.* 1994). Subsequently, it has been shown that the 86Y mutation in *mdr1* increases the CQ IC_{50} in *P. falciparum* parasites harbouring *crt*-76T (Babiker *et al.* 2001). In Kenya, the triple mutant *mdr1* haplotypes YYY (at codon 86, 184 and 1246) reached frequencies greater than 55% pre-

ACT introduction but has since not been detected in the population, similar to the *crt* CQR haplotypes (Okombo *et al.* 2014).

1.9.2.2 Sulfadoxine-Pyrimethamine (SP) resistance

Sulfadoxine and pyrimethamine belong to a class of antimalarial drugs called antifolates because they target the parasites folate pathway (Ross and Fidock 2019). SP replaced CQ as CQR was associated with a significant increase in mortality (Trape *et al.* 1998). Soon after, resistance to SP emerged, spread and was also associated with an increase in malaria-related mortality (Korenromp *et al.* 2003). For this reason, SP is no longer used for the treatment of clinical malaria, however, it continues to be used as prophylaxis and is administered routinely as intermittent preventive treatment in pregnant women (IPTp) (WHO 2014b). The antifolate effects of SP are derived from the inhibition of two enzymes in the folate synthesis pathway, dihydropteroate synthase (*dhps*) (Zhang and Meshnick 1991) by sulfadoxine and dihydrofolate reductase (*dhfr*) by pyrimethamine (Olliaro and Yuthavong 1999), eventually inhibiting nuclear division.

The *dhfr* gene encodes a bifunctional protein, consisting of the *dhfr* domain in the first 231 amino acids and the TS domain in the last 288 residues, separated by 89 residues of the inter-junction region between them (Bzik *et al.* 1987). Multiple mutations in *dhfr* at codons 50, 51, 59, 108, and 164 have been demonstrated to be the major determinant of pyrimethamine resistance. However, the mutation at codon 108 appears to be the most essential for *in vitro* pyrimethamine resistance while additional mutations synergistically increase the levels of resistance. For example, 108N mutation increases the IC_{50} of pyrimethamine by up to 50 fold above that of the wild-type S108, while the double mutants S108N+N51I or S108N+C59R, increase resistance by up to 16 fold over the 108N single mutant (Peterson, Walliker and Wellems 1988; Foote, Galatis and Cowman 1990). Other studies have shown that triple mutants (N51I+C59R+S108N and C59R+S108N+I164R) demonstrate higher IC_{50} to pyrimethamine compared to double mutants (Peterson, Walliker and Wellems 1988; Foote, Galatis and Cowman 1990; Basco *et al.* 1995) while others find that the quadruple mutant (N51I+C59R+S108N+I164R) has the highest IC_{50} to pyrimethamine so far (Sirawaraporn *et al.* 1997).

The *dhps* gene is the target of sulfadoxine and is encoded by a single copy gene and *dhps* mutations confer resistance to sulfadoxine by decreasing affinity for the enzyme (Triglia *et al.* 1997). To date, five mutations have been associated with sulfadoxine resistance, including 436, 437, 540, 581, and 613 in a stepwise manner. In an earlier study (Triglia *et al.* 1998), the mutation at codon A437G was found to confer a 4.8-fold increase in the IC_{50} to sulfadoxine over wild-type parasites. However, an additional mutation at codon S436A did not significantly increase the IC_{50} to sulfadoxine at either codon S436A or A581G. On the other hand, triple *dhps* mutants conferred a substantial increase in sulfadoxine resistance, a 9.8-fold increase over wild-type (S436A+A437G+K540E) and a 24-fold increase over wild-type (S436A+A437G+A613S). The mechanism by which these mutations confer resistance is not yet fully understood, but one study suggested that the mutations may result in lowered affinity of sulfadoxine for the active site (de Beer, Louw and Joubert 2006).

Similar to other sites in Africa, Kenya has a high frequency of SP resistance markers (Mwai *et al.* 2009b; Okombo *et al.* 2014) even though SP was withdrawn as first-line treatment and highlights the need for ongoing surveillance since SP is still recommended for IPTp.

1.9.2.3 Artemisinin combination therapy resistance

The WHO recommends ACTs as first-line treatment for *falciparum* malaria, following widespread resistance to former first-line treatments (CQ and SP) (WHO 2015). Sufficient parasitological and clinical therapy by ACT drugs is dependent on the rapid reduction in parasite biomass by the short-acting but potent artemisinin component followed by the elimination of residual parasites by the longer-acting partner drug (White, Van Vugt and Ezzet 1999). Various combinations of ACTs exist, but artemether-lumefantrine (AL) and artesunate-amodiaquine (AL+AQ) are the first-line treatment policies used in most SSA countries, with some countries adding dihydroartemisinin-piperaquine (DHA+PPQ) as second-line treatment (WHO 2018b). The success of ACT use is evident as substituting failing treatments of chloroquine (CQ) and sulfadoxine-pyrimethamine (SP) with ACTs has been associated with reduced malaria-related morbidity and mortality (Bhatt *et al.* 2015; Gething *et al.* 2016). However, the emergence and spread of artemisinin resistance in SE Asia (Noedl *et al.* 2008; Dondorp *et al.* 2009; Ashley *et al.* 2014) is of great significance as in the past, resistance to CQ

and SP were imported to Africa from Asia (Wootton *et al.* 2002; Roper *et al.* 2004; Mita *et al.* 2011). Nonetheless, some studies have shown that some *dhps*, *dhfr* and *mdr1* alleles emerged independently in Africa based on typing of microsatellite markers (Alam *et al.* 2011).

1.9.2.3.1 Artemisinin

The mode of action of artemisinin is believed to involve iron-catalysed scission of its endoperoxide bridge that leads to widespread alkylation and oxidative damage inside the asexual blood stage parasite (Wang *et al.* 2015a; Ismail *et al.* 2016). The primary source of iron that is required to activate artemisinin is mainly derived from parasite-mediated digestion of host haemoglobin digestion (Xie *et al.* 2016). The clinical derivatives of artemisinin (dihydroartemisinin, artemether and artesunate) have the potential to reduce the biomass of a drug-sensitive infection by up to 10,000-fold every 48-hour asexual blood-state cycle. However, artemisinin's half-life in plasma is very short, only 1–2 h (White 1997, 1999). For this reason, artemisinin derivatives are paired in artemisinin combination therapy (ACT) with a longer-acting partner drug including, lumefantrine, amodiaquine, piperaquine, SP, mefloquine and pyronaridine (Ross and Fidock 2019). This, however, means that after artemisinin is eliminated in the blood, there is an increased risk of partner-drug resistance since the partner drug remains as a monotherapy.

The earliest reports of delayed response to artemisinins in *P. falciparum* parasites were from western Cambodia in 2008 to 2009, leading to the identification of the hallmark of artemisinin resistance, slow parasite clearance *in vivo* (Noedl *et al.* 2008; Dondorp *et al.* 2009). By 2014, slow parasite clearance had become common across countries in the Greater Mekong Subregion, that includes Cambodia, China, Laos, Myanmar, Thailand and Viet Nam (Ashley *et al.* 2014). The development of *in vitro* ring-stage survival assays (RSAs) with synchronized *P. falciparum* rings exposed to dihydroartemisinin (DHA) treatment at 6 hour intervals, led to a clearer definition of artemisinin resistance (Witkowski *et al.* 2013). These assays revealed that the 0-3 hour rings were less susceptible to dihydroartemisinin in artemisinin-resistant parasites, since artemisinins act on young ring-stage parasites (White 2017) and were therefore used as an *in vitro* correlate for artemisinin resistance. Later studies showed that a high proportion of Cambodian samples carried single mutations in the *P. falciparum kelch13*

protein (*Pfk13*) that correlated with parasite survival rates *in vitro* and day 3 parasitemia in therapeutic efficacy studies and this led to the identification of *k13* as the molecular marker for artemisinin resistance (Ariey *et al.* 2014).

The *Pfk13* gene encodes a protein containing a conserved N-terminal domain followed by a BTB/POZ and finally a 6-blade propeller domain in the C-terminus (codons 441–725) (Ariey *et al.* 2014). Mutations in the propeller domain have been associated with artemisinin resistance, but few have been validated to date including F446I, N458Y, M476I, Y493H, R539T, I543T, P553L, R561H and C580Y (Ariey *et al.* 2014; Ashley *et al.* 2014). *k13* mutations associated with artemisinin resistance arose independently along the Myanmar-Thai (Takala-Harrison *et al.* 2015) and Cambodia-Thai (Miotto *et al.* 2013) border regions. Consequently, whole-genome studies of artemisinin-resistant parasites in SE Asia were shown to accumulate artemisinin resistance predisposing mutations at other genetic loci, namely codon V127M in apicoplast ribosomal protein S10 precursor (*arps10*), codon I356T in *crt*, codon D193Y in ferredoxin (*fd*) and codon T484I in multidrug resistance (*mdr*) protein 2 (Miotto *et al.* 2015).

One mutation, however, C580Y, has gone on to be the most dominant and studies have shown that it first arose in western Cambodia before it spread to Thailand and Laos (Imwong *et al.* 2017). Worryingly, it appears that some parasites carrying this mutation have gone on to acquire partner drug-resistance to piperaquine, characterized by amplification of a gene called *plasmepsin-2*, and have spread from Cambodia to Thailand and Laos (Imwong *et al.* 2017). As a consequence, Cambodia has adopted AS-MQ as the first-line treatment due to widespread DHA-PPQ, while Thailand and Vietnam are reviewing current policy and procedures (WHO 2018b). Numerous subgroups of highly related parasites have also emerged within this co-lineage of multi-drug resistant parasites with diverse geographical distributions (Hamilton *et al.* 2019), hence a 'firewall' of containing artemisinin resistance may not work and calls for the elimination of this increasingly drug-resistant parasite co-lineage.

Pfk13 mutations associated with artemisinin resistance have not been observed in Africa (Kamau *et al.* 2015; Taylor *et al.* 2015; Ménard *et al.* 2016; WWARN K13 Genotype-Phenotype Study Group 2019), although a recent study reported the Y493H, P553L, R561H and C580Y

mutations in low frequencies (<0.1%) (MalariaGEN Plasmodium falciparum Community Project 2016). Also, though lacking the validated *k13* mutations associated with artemisinin resistance, parasites with high artemisinin survival rates have been identified in Equatorial Guinea and Uganda (Lu *et al.* 2017; Ikeda *et al.* 2018). In contrast, some studies have reported artemisinin resistance in isolates that did not harbour *k13* mutations, termed "discordant" parasites. In a study conducted by Mukherjee et al. (2017) using the samples collected from Cambodia, the authors reported a significant positive correlation between the *k13*-C580Y mutation and *in vivo* parasite clearance half-life values. However, four of the thirty-six isolates were reported to have high *in vitro* RSA survival rates (>1%) compared to artemisinin sensitive isolates, yet they lacked *k13* mutations. While this small data set was underpowered to identify novel mutations using a genome-wide strategy, the authors used a candidate-gene approach and identified SNPs in other genes including *dhps*, *dhfr*, *mdr2* and phosphatidylinositol-4-phosphate 5-kinase (PIP5K). In another study, Demas et al. (2018) carried out long-term *in vitro* selection in two recently isolated Senegalese parasites with increasing concentrations of DHA over a 4-year period. The authors isolated two parasite clones, one from each original isolate, that exhibited enhanced survival to DHA in the RSA and these clones harboured up to 10 SNPs that were selected for, *in vitro*. Notably, the only candidate drug resistance gene to be identified in both clones was the actin-binding protein coronin (*Pf*coronin) with polymorphisms at codons R100K, E107V and G50E in the absence of *k13* mutations. These two studies present evidence to show that other genes other than *k13* may be involved in conferring artemisinin resistance. Therefore, *k13* alone might not be enough to track artemisinin resistance especially in regions such as in SSA where resistance is yet to be reported.

A mutation (S69Stop) in the cysteine proteinase *falcipain-2a* gene has previously been selected for after artemisinin *in vitro* selection pressure, using a parasite isolate from Africa. Previous work has shown that artemisinin activity is dependent on haemoglobin uptake and digestion. *Pffalcipain-2a* encodes a cysteine protease that is involved in this pathway and its deletion has been significantly associated with decreased artemisinin sensitivity (Klonis *et al.* 2011). Therefore, it appears that a mutation in *Pffalcipain-2a* may interfere with its function,

thereby modulating artemisinin resistance. A clinical trial conducted in Western Kenya, revealed that 2-3 days of ACT treatment selected for parasites with either the 160N or 160T alleles in the AP-2 complex subunit mu gene (*Pfap2-mu*, S160N/T) and the allele 1528D in the ubiquitin carboxyl-terminal hydrolase 1 gene (*Pfubp-1,* E1528D) (Henriques *et al.* 2014). Both *Pfap2-mu* and *Pfubp-1* are likely to also be involved in haemoglobin digestion (Birnbaum *et al.* 2020) and just like *Pfap2-mu*, mutations in these two genes are likely to interfere with their functions and render parasites less sensitive to artemisinin.

1.9.2.3.2 Lumefantrine

Some of the partner drugs used with artemisinin include lumefantrine, a drug that belongs to the aryl-amino alcohols class of antimalarials. This combination, artemether-lumefantrine (AL) is the most widely used antimalarial in endemic countries, accounting for approximately 75% of all procured ACTs (UNITAID 2018). Lumefantrine's mode of action is not well understood and its main target remains unsolved, but it is thought to interfere with haem detoxification (Combrinck *et al.* 2013) or to directly inhibit *mdr1* (Martin, Shafik and Richards 2018). Several recent studies have associated some polymorphisms associated with lumefantrine pressure and warrant further investigation. AL has been shown to select for the *Pfcrt*-K76 allele and the *Pfmdr1* NFD haplotype (N86/184F/D1246) (Sisowath *et al.* 2005, 2009; Humphreys *et al.* 2007; Henriques *et al.* 2014). Furthermore, the *Pfmdr1*-NFD haplotype and an increase in *Pfmdr1* copy number have been linked with reduced susceptibility to lumefantrine (LM). Recently, a temporal increase in the frequency of the K65 allele (K65Q) in the cysteine desulfurase (*nfs*) gene was observed in The Gambia, six years after the introduction of ACTs. Moreover, the IC_{50} values for LM were significantly higher in *nfs*-K65 wild-type field isolates when compared to the mutant 65Q isolates (Amambua-Ngwa *et al.* 2018). Exactly how *Pfcrt*, *Pfmdr1* and *Pfnfs* modulate lumefantrine resistance, is not known.

1.9.2.3.3 Amodiaquine

Artesunate-amodiaquine (ASAQ) is the second most used ACT combination drug in malaria endemic settings. Amodiaquine belongs to the 4-Aminoquinolines class of antimalarials and

in 2017, ASAQ accounted for greater than 20 % of all ACTs (UNITAID 2018). Amodiaquine (AQ) has also been adopted for seasonal malaria chemoprevention in the Sahel region, in combination with SP (Coldiron, Von Seidlein and Grais 2017). The mode of action of AQ is similar to CQ, as it accumulates in the parasite's digestive vacuole where it binds the toxic haem, preventing its detoxication into the inert form hemozoin (Kaur *et al.* 2010). AQ resistance has been associated with point mutations in the *Pfcrt* and *Pfmdr1* genes, and the same mutations as those for CQ resistance (*Pfcrt* 76T and *Pfmdr1* 86Y) have been shown to be the main determinant of decrease susceptibility to AQ *in vitro* and *in vivo* (Duraisingh *et al.* 1997; Holmgren *et al.* 2006; Yeka *et al.* 2016).

1.9.2.3.4 Piperaquine

The combination of dihydroartemisinin and piperaquine (DP) is an e in effective drug most malaria endemic settings (Achan *et al.* 2013) and has been used in some countries in Southeast Asia and Africa (WHO 2018a). This combination is increasingly being considered for malaria prevention in pregnancy (Kakuru *et al.* 2016) as well as mass drug administration (MDA) in near-elimination settings in Africa (Eisele *et al.* 2016; Mwesigwa *et al.* 2019). The mode of action of PQ is yet to be understood, but it has been linked to the inhibition of haemoglobin degradation pathway (Dhingra *et al.* 2017). DP is failing in the sub-Mekong region due to resistance to piperaquine (PQ) and decreased susceptibility to artemisinin derivatives (Amaratunga *et al.* 2016; Amato *et al.* 2017). More recently, PQ resistance has been associated with copy number variation in *Pfplasmepsin-II* and *III* (Witkowski *et al.* 2017) and point mutations in *Pfcrt* (Agrawal *et al.* 2017; Dhingra *et al.* 2017; Ross *et al.* 2018). *in vivo* assays are limited in the assessment of phenotypic resistance to PQ, however, a PQ survival assay (PSA) has been established to assess changes in parasite strains' susceptibility to PQ (Witkowski *et al.* 2017).

1.9.2.3.5 Sulfadoxine Pyrimethamine

Notwithstanding widespread resistance to SP in most malaria endemic settings, the combination artesunate and SP (ASSP) is still used in a some countries for malaria treatment (WHO 2018a) and in IPTp as mentioned earlier. SP is also used in combination with

amodiaquine for seasonal malaria chemoprevention (SMC) (Coldiron, Von Seidlein and Grais 2017). The mechanism of resistance to SP has been described earlier in this chapter.

1.9.2.3.6 Mefloquine

The combination of artesunate and mefloquine (ASMQ) was the first ACT introduced in Southeast Asia to halt resistance to mefloquine (MQ) monotherapy in the early 1990s (Trenholme 1993). Noteworthy, the use of this combination has historically been restricted to Southeast Asia, the Pacific region and South America, but was not used in Africa (WHO 2019). With widespread resistance to MQ and decreased susceptibility to artemisinin derivatives, ASMQ is now used only in a few countries (WHO 2019). The mode of action of MQ is still unclear, but the drug may inhibit haem detoxification (Fitch 2004; Combrinck *et al.* 2013) or directly inhibit pfmdr1 (Cowman, Galatis and Thompson 1994). MQ resistance has also been linked to copy number variations in *Pfmdr1* (Price *et al.* 2004) as well as polymorphisms in *Pfmrp1* and *Pfmrp2* (Veiga *et al.* 2011, 2014; Woodland *et al.* 2018). More recently, the cytoplasmic ribosome (*Pf*80S) of the asexual blood-stage parasite has been suggested as the main target of MQ (Wong *et al.* 2017).

1.9.2.3.7 Pyronaridine

The combination of artesunate and pyronaridine (ASPY) is not yet recommended by the WHO, but is the only ACT indicated for the blood stage treatment of both *P. falciparum* and *P. vivax* (Medicines for Malaria Venture). The combination has shown high efficacy in Africa and was non-inferior to AL and DP (Sagara *et al.* 2016, 2018), and in Eastern Cambodia (Leang *et al.* 2019). The mode of action of PY is not well known, however, it is thought to interfere with the hemozoin formation (Auparakkitanon *et al.* 2006). Resistance to PY has not been reported, however, *ex vivo* assessment has shown an association between decreased susceptibility to PY and the *Pfcrt*-76T mutation (Madamet *et al.* 2016). The findings highlighted herein, underscore the need for continued surveillance of artemisinin and partner drug resistance markers in Africa, where the burden of malaria is the highest, should it emerge independently or spread from SEA.

1.9.3 Chronic asymptomatic infections

Asymptomatic *falciparum* infections are prevalent in malaria-endemic SSA, reaching over 90% in children (Okell *et al.* 2009). A large proportion of asymptomatic infections are associated with sub-microscopic infections that are not detected in routine microscopy tests and hence go untreated, which can result in a major parasite reservoir that can fuel malaria transmission and negatively impact on malaria elimination efforts (Alves *et al.* 2005). Accordingly, the use of sensitive nucleic acid-based techniques (PCR) has shown that the prevalence of asymptomatic infections is approximately two-fold compared to those identified with conventional diagnostics, such as microscopy (Okell *et al.* 2009) and RDT (Wu *et al.* 2015). Due to high herd immunity and premunition in the population, asymptomatic infections are expected to be common in high malaria transmission areas. However, sensitive PCR-based diagnostics have revealed that asymptomatic infections are also common in low-endemic areas despite residents having less NAI compared to those living in areas of high malaria transmission (Okell *et al.* 2012). The interest in asymptomatic infections has focussed both on their impact on health as well their role in malaria transmission and elimination as highlighted below.

One major obstacle in the study of asymptomatic malaria is the lack of standard definition. Infected individuals may be in a pre-symptomatic period with parasitaemia and end up presenting with symptoms of malaria at a later date (Coura, Suárez-Mutis and Ladeia-Andrade 2006). Alternatively, studies that do not incorporate an individual's clinical history may fail to capture those that may have experienced symptoms briefly and took medication that suppressed parasitaemia and symptoms. The most widely-used criteria for defining asymptomatic malaria is the presence of *P. falciparum* parasites in peripheral thick blood smears, with no symptoms indicative of malaria (Abdel-Latif *et al.* 2003; Males, Gaye and Garcia 2008). Some studies have included other criteria, such as longitudinal follow-up. Longitudinal follow-up is important especially for identifying infections that appear asymptomatic at time of diagnosis, but may subsequently become symptomatic (Krieger *et al.* 2002; Leoratti *et al.* 2008; Harris *et al.* 2010). While species-specific PCR may not always be available for use in the field, the higher sensitivity it offers compared to the standard

microscopy has been shown to detect an even greater number of asymptomatic infections, underscoring just how many asymptomatic infections may be misdiagnosed as negative (Bousema *et al.* 2014). The inclusion of genetic disorders such sickle cell trait, that are known to protect against malaria, has also been explored so as to exclude such individuals who may be infected but do not develop symptoms of malaria (Touré *et al.* 2006). The varying definitions of asymptomatic malaria may pose challenges in comparing the burden of asymptomatic infections across different geographical settings and quantifying the impact of control or elimination programs that target these infections (Lindblade *et al.* 2013).

Asymptomatic infections are thought to be as a result of partial immunity which controls but does not eliminate the infection. For this reason, these infections are sometimes viewed as beneficial to the individual as they help to maintain a state of premunition, thereby reducing the risk of severe disease (Doolan, Dobaño and Baird 2009). However, there is growing evidence that asymptomatic infections can be detrimental to the host, with serious health, developmental, and productivity consequences, leading to some describing them as "chronic" infections that require curative treatment (Chen *et al.* 2016). Some studies have associated chronic asymptomatic infection with elevation of C-reactive protein (CRP), a biomarker of inflammation (de Mast *et al.* 2015; Peto *et al.* 2016), lower platelet counts and haemoglobin levels as well as higher levels of von Willebrand factor and platelet factor-4, markers of endothelial and platelet activation, respectively (de Mast *et al.* 2015), low birth weight and premature births in pregnancy (Cottrell *et al.* 2015), cognitive impairment (Nankabirwa *et al.* 2013), malnutrition (Maketa *et al.* 2015) and anaemia (Matangila *et al.* 2014; Maketa *et al.* 2015; Sifft *et al.* 2016).

Regarding the impact of asymptomatic infections on transmission, the membrane feeding assay has been the tool used to assess the infectivity of asymptomatic individuals to mosquitoes. For example, one study conducted in Western Thailand assessed the infectivity of asymptomatic individuals harbouring *P. falciparum*, *P. vivax*, *P. malariae* infections and found that 10% (7/70), 13% (7/52), and 0% (0/6) infected mosquitoes, respectively (Coleman *et al.* 2004). In another study conducted in Western Kenya, residual parasitemia detected by PCR after ACT administration was associated with high gametocyte carriage in children and

such children had an increased capability of infecting mosquitoes (Beshir *et al.* 2013). These studies highlight the significant role played by asymptomatic infections in fuelling malaria transmission.

There is conflicting literature on the role of asymptomatic infections on the risk of developing febrile malaria with some studies reporting that asymptomatic infections are associated with increased risk of febrile malaria (Henning *et al.* 2004; Njama-Meya, Kamya and Dorsey 2004; Le Port *et al.* 2008; Males, Gaye and Garcia 2008) while other studies have reported a reduced risk (Henning *et al.* 2004; Males, Gaye and Garcia 2008; Portugal *et al.* 2017; Buchwald *et al.* 2018). Overall, in the high transmission settings, there is a reduced risk of developing febrile malaria and this could be stratified by age such that older children (>5 years) were at a reduced risk compared to younger children (≤5 years), while the moderate to low transmission settings showed an increased risk of developing febrile malaria regardless of age. While these earlier studies were conducted in different geographical settings and with varying sample sizes and age groups, they emphasise that asymptomatic infections predict the risk of subsequent febrile malaria infections. Still, how other factors such as age and transmission intensity impact on this risk is yet to be understood.

The indication that premunition is one of the mechanisms by which asymptomatic individuals develop partial immunity to malaria would suggest that curing or preventing asymptomatic infections would then make these individuals more susceptible to clinical attacks. Various studies have evaluated this and it appears that there are conflicting reports with some studies revealing that treatment of asymptomatic infections was associated with increased risk of clinical attacks (Owusu-Agyei *et al.* 2002; Liljander *et al.* 2010, 2011; Ouédraogo *et al.* 2010; Tiono *et al.* 2013; Larsen *et al.* 2015) while other studies found no effect (Portugal *et al.* 2017). Given the conflicting literature regarding the role of asymptomatic infections on developing febrile malaria infections, further studies are needed to clarify this.

As already highlighted, age and malaria transmission intensity modify the risk of developing subsequent febrile episodes, complexity of infection (COI) has also been reported to be associated with the risk of developing febrile episodes. In line with this, previous reports have shown that the dynamics of asymptomatic infections are complex and reveal that parasite

genetic diversity is greater in high-transmission settings compared to low-transmission settings (Magesa *et al.* 2002; Oyedeji *et al.* 2013; Adjah *et al.* 2018). Additionally, asymptomatic malaria infections have been found to harbour more complex infections compared to febrile malaria infections (Beck *et al.* 1997; Magesa *et al.* 2002) and exhibit a frequent turnover of clones (Daubersies *et al.* 1996; Nkhoma *et al.* 2018). Moreover, this high genetic diversity within asymptomatic infections has been associated with a reduced risk of developing febrile malaria episodes (Al-Yaman *et al.* 1997; Bereczky *et al.* 2007; Sondén *et al.* 2015), although, the inverse has also been observed (Felger *et al.* 1999; Branch *et al.* 2001; Ofosu-Okyere *et al.* 2001; Henning *et al.* 2004). Conversely, lower COI in febrile malaria episodes could be explained by the fact that most fever episodes appear to be associated with parasitaemia peaks (Cox *et al.* 1994) and this may result from the expansion of just a few or even single genotypes. Therefore, this could lead to sampling of only a few clones leading to minority clones remaining undetectable. Fever is also known to have an antiparasitic effect and this may result in the elimination of some genotypes from the initial parasite pool (Kwiatkowski 1991). One study conducted in Sudan, however, has previously reported that the proportion of infections with more than 1 clone was greater in febrile malaria infections (51%) than in asymptomatic infection (38%) (Roper *et al.* 1998). The authors hypothesized that because febrile malaria infections had higher parasite densities, sampling greater number of parasites increased the likelihood of detecting mixed infections. Another study conducted in Malawi, it was shown that polyclonal infections were more common in febrile malaria infections (302/477, 63.3%) compared to asymptomatic infections (180/585, 30.8%, $p < 0.001$) (Earland *et al.* 2019). However, this may have been as a result of the study's inability to genotype samples with low parasite densities which were mainly asymptomatic samples. The different findings from multiple studies regarding parasite diversity in asymptomatic vs. febrile infections warrants the use of a more sensitive tool coupled with intermittent sampling so as to efficiently capture of all circulating parasite clones.

Most malaria interventions target the febrile population because they present with clinical symptoms of malaria. However, the high prevalence of asymptomatic infections necessitates the inclusion of asymptomatic parasite carriers in drug treatment interventions (Okell *et al.*

2011; Lubis *et al.* 2017), though, further studies are needed to clarify the impact of treating asymptomatic infections on the risk of subsequent febrile malaria infections. This may have a larger impact in preventing transmission as opposed to interventions that only target febrile cases, as asymptomatic infections may have long duration. As research priorities shift from malaria control to malaria elimination, it is important to tailor interventions that target both febrile and asymptomatic infections as this will not only reduce the malaria burden but also interrupt malaria transmission. Consequently, a deeper understanding of asymptomatic infections is needed.

1.9.4 Parasite genetic diversity

There is great genetic diversity within *P. falciparum* and individuals living in malaria-endemic regions are frequently and simultaneously infected by several distinct parasite clones (Juliano *et al.* 2010). Harbouring multiple genetically distinct parasites, termed “complexity of infection (COI)”, is one of the reasons for the slow acquisition of NAI leading to the view that humans must encounter several diverse parasite clones before they develop effective immunity to malaria (Doolan, Dobaño and Baird 2009). Diversity in *P. falciparum* is generated for example during fertilisation between gametes and zygote formation in the vector and novel parasite genotypes can arise through chromosome reassortment at meiosis and recombination (Sinden 2009). Consequently, single genetically distinct clones are then transmitted through multiple mosquito bites or via single mosquito bites with multiple clones, especially in areas of higher transmission. Likewise, surface antigens (targets of NAI) expressed by the parasite have also been shown to be highly diverse through the acquisition of extensive point mutations, but the precise targets that would correlate with immunity are unknown at present (Doolan, Dobaño and Baird 2009).

This extensive parasite diversity, as well as interactions between genetically distinct malaria parasites within a single host, has been shown to impact on several outcomes including malaria transmission, the evolution of parasite virulence, disease progression, antimalarial drug resistance, immunity and promote "vaccine-resistant" clones. In a rodent model, de Roode et al. (2005) showed that there was a strong relationship between parasite virulence and competitive ability, such that the more virulent clones had a competitive advantage in

mixed-clone infections. Moreover, there was a direct correlation between the clone composition of the parasite populations in mosquitoes and that of the blood-stage parasite population. Bell et al. (2006) determined the pairwise competitiveness of genetically divergent *P. chabaudi* clones with a wide range of virulence in their rodent host. The study demonstrated that virulent clones had a competitive advantage in the acute phase of mixed infections. This led the authors to conclude that such models show how virulence is selected for, indicating the need for interventions that reduce the incidence of mixed infections. May et al. (2000) determined the impact of polyclonal infections on haematological parameters in clinically healthy children in Nigeria and revealed that 64% of the 228 children recruited were anaemic. The prevalence of anaemia was dependent on the number of *Plasmodium* clones detected ($P < 0.0001$) and the prevalence of anaemia increased gradually with the complexity ($P < 0.003$) as well as with the extent of *P. falciparum* parasitaemia ($P < 0.0001$). Wargo et al. (2007) studied competitive release in acute infections containing drug-susceptible and resistant clones following pyrimethamine treatment in a rodent model. They showed that there was an expansion of resistant clones and enhanced gametocyte densities after treatment. The number of resistant clones increased beyond that achieved when a competitor had never been present, thus this competitive release substantially increased the fitness advantage of resistant clones. Within-host parasite diversity has been found to correlate with immunity in a study conducted in children aged between 1-16 years in Tanzania (Rono *et al.* 2013). By measuring the number of merozoite antigens to which participants responded to, Rono et al. (2013) showed that the breadth of antibody responses was positively correlated to the number of infecting clones. Likewise, the breadth of antibody responses and the number of infecting clones were associated with a lower risk of malaria. Lastly, the RTS,S/AS01 vaccine (trade name, Mosquirix) based on the circumsporozoite protein (*csp*), expressed during the invasion of liver cells, is the foremost malaria vaccine to be licensed for use by the European Medicines Agency. However, the RTS,S/AS01 has been associated with only 18–36% vaccine efficacy against clinical malaria of (RTS-S Clinical Trials Partnership 2015). One of the reasons attributed to the limited efficacy of RTS,S/AS01 is based on the fact that *csp* is a globally diverse gene (Barry *et al.* 2009). Consequently, cumulative

vaccine efficacy of RTS,S/AS01 in Phase III clinical trials was reduced from 50.3% for parasites with a perfect *csp* C-terminal sequence match to only 33.4% for parasites with any amino acid mismatch in this region (Neafsey *et al.* 2015).

The studies presented here highlight how *P. falciparum* has evolved to harness genetic diversity to its advantage. Accordingly, a better understanding of this diversity is crucial in informing malaria elimination efforts.

1.10 Genetic epidemiology of *P. falciparum*

SNP genotyping methods have previously been applied to analyze parasite genetics, but this has been limited to research settings. Thus, their utility when incorporated in National Malaria Control Programmes has not yet been proven, but may help accelerate progress towards malaria elimination as discussed below.

SNP genotyping methods have been used before to detect, track and provide surveillance of drug-resistance in *P. falciparum*, a feat that was made possible by the identification of drug-resistance markers. Notably, *crt*, *mdr1, dhps, dhfr* and *k13* were defined as molecular markers of resistance to CQ, SP and artemisinin, respectively (Peterson, Walliker and Wellems 1988; Foote *et al.* 1989; Wilson *et al.* 1989; Foote, Galatis and Cowman 1990; Triglia *et al.* 1998; Fidock *et al.* 2000; Ariey *et al.* 2014). Such information has been used to study natural populations of *P. falciparum* across diverse geographical regions and enabled studies to carry out genetic epidemiology of drug-resistance markers to CQ (Wootton *et al.* 2002), SP (Roper *et al.* 2004; Mita *et al.* 2011) and more recently, artemisinin (Ashley *et al.* 2014; MalariaGEN Plasmodium falciparum Community Project 2016; Ménard *et al.* 2016). In the wake of artemisinin resistance, drug-resistance surveillance using molecular markers offers a scalable tool to detect and track resistance to artemisinins especially in regions such as SSA, where it is yet to be reported.

SNP genotyping methods have also been used to define within-host parasite diversity and transmission patterns in populations. Such information can be used to assess the effects of malaria interventions, such as the use of IRS, ITNS and ACTs, as they are expected to reduce parasite diversity. The prevalence of polyclonal infections in a population can be a good

indicator of transmission intensity since a positive correlation between the rate of polyclonal infections and annual parasite incidence has been observed (Fola *et al.* 2017). Some studies have reported that a reduced risk of clinical malaria was associated with polyclonal infections (Al-Yaman *et al.* 1997; Bereczky *et al.* 2007), while other studies reported that infections with one clone were associated with severe malaria compared to infections with more than one clone (Durand *et al.* 2008). In a study conducted in Senegal, Daniels et al. (2015) showed that parasite diversity decreased and later increased in tandem with reductions in malaria transmission intensity due to malaria intervention efforts and later a rebound in malaria cases, respectively. Following malaria intervention efforts in Malawi, Sisya et al. (2015) reported that parasite diversity in a single location (Blantyre) remained high, suggesting that only subtle gains, if any, were made in reducing malaria transmission.

SNP genotyping has been used to identify malaria foci linking episodes based on parasite genetic relatedness and hence such information can aid in the identification of malaria foci that can inform targeted malaria intervention strategies. Following reports in Madagascar that there existed localities with transmission and prevalence rates up to 10-fold higher than the national average, Rice et al. (2016) identified the Makira region of Madagascar as a hotspot. In this region, a high level of genetic diversity and a high frequency of infections with multiple *P. falciparum* clones were observed, providing a pattern consistent with the high and stable transmission. In a study conducted in Greece, Spanakos et al. (2018) used microsatellites markers to attribute an outbreak of *P. vivax* malaria to specific villages with foci revived by imported malaria via migrant agricultural workers.

Lastly, SNP genotyping has been used to determine the connectivity of parasite populations. In turn, parasite populations are linked across various geographic regions, facilitating a quantitative assessment of how much of transmission in one region is due to contributions from transmission in another. This information can be used to tailor malaria interventions that target imported cases of malaria. In a study conducted in Papua New Guinea, Fola et al. (2018) used microsatellite markers to investigate population structure in eight geographically and ecologically distinct regions of the country. Parasite populations were found to be increasingly divergent with geographic distance and that one region acted as a “sink”, where

infections arrived from all other populations with limited outbound flow. In another study conducted in Western Kenya, Omedo et al. (2017) found no barriers to the spatial movement of parasites in the region because the parasites exhibited no population structure. Markedly, this study highlighted that the high level of parasite mixing suggested that a geographically restricted "one-off" targeted intervention may not be currently effective in Western Kenya, due to the high parasite movement that is likely to lead to re-introduction of infections from surrounding regions.

1.11 Molecular approaches to determine parasite diversity

1.11.1 Size polymorphic marker and microsatellite markers

The traditional method for characterizing *P. falciparum* diversity uses nested-polymerase chain reaction (nested PCR) with gel electrophoresis to detect polymorphisms in merozoite surface protein 1 and 2 (*msp1* and *msp2*) and glutamine-rich protein (glurp) (Viriyakosol *et al.* 1995). Microsatellite markers are highly polymorphic with tandem repeats of between 2-6 base pairs (bp) that are not under selection, show no phenotypic consequences and are extensively distributed throughout the parasite genome (Anderson *et al.* 1999). These markers are readily amplified by PCR to estimate COI and typically, COI is determined as the number of alleles detected in a gel and several markers can be combined to resolve bias. However, the reliance on gel electrophoresis to determine allele sizes for *msp1*, *msp2*, *glurp* and microsatellite markers is limited in discriminating alleles of similar sizes (those with size difference less than 20 bp). Moreover, technical problems such as non-specific amplification (de Valk *et al.* 2009) and stutter peaks (de Valk, Meis and Klaassen 2007) are some of the obstacles to using these markers to assess COI. Consequently, these markers underestimate diversity, are insensitive to low-abundant variants and are not quantitative for relative proportions of circulating parasite clones.

1.11.2 Sanger sequencing

Sanger sequencing is the di-deoxy chain termination method for DNA sequencing by synthesis pioneered by Fred Sanger. One strand of the double-stranded DNA is used as a template to be sequenced, and chemically-modified, fluorescently-labelled di-deoxynucleotides (ddNTPs)

are incorporated. Unlabelled dNTPs are present in excess in the elongation of DNA strands, so the incorporation of ddNTPs into the growing DNA strand, preventing further elongation, is a random event. This process generates DNA fragments of different sizes ending in labelled ddNTPs. These fragments are separated according to their sizes on a gel where the resultant bands, corresponding to DNA fragments, can be visualized by an imaging system including X-ray and UV light (Sanger and Coulson 1975) or in current practice, detection of specific fluorophores for each nucleotide. The major limitation of Sanger sequencing is that it cannot always identify multiclonal infections as well as being time and cost-prohibitive, especially with large sample sizes. Nonetheless, Sanger sequencing remains the gold standard for confirming DNA sequences due to the stability of the technology and it is still broadly used for targeted sequencing analysis in research and clinical laboratories (Alekseyev *et al.* 2018).

1.11.3 High-throughput (HT) SNP genotyping

Numerous panels of genome-wide SNP markers have been identified and protocols for typing developed. One panel includes 24 *P. falciparum* SNPs selected from over 112,000 SNPs of 18 parasite genomes (Daniels *et al.* 2008) whose greatest power is to discriminate among clones compared to the traditional *msp-1* and *msp-2*. HT-SNP genotyping can also be done on the Sequenom MassARRAY iPLEX platform, which can multiplex up to 40 SNPs per single reaction. The genotype of a parasite is determined by the SNP allelic intensity ratios, ranging between 0 and 1. The SNP allelic intensity ratio values nearing 0 and 1 indicate a single parasite genotype infection, while intermediate values indicate mixed parasite genotype infections (Gabriel, Ziaugra and Tabbaa 2009). The TaqMan assay is another HT-SNP genotyping method that PCR-based. In this assay, the region flanking a SNP is amplified in the presence of two allele-specific fluorescent probes that allow the detection of both alleles in a single tube. Compared to msp1/msp2/glurp genotyping, genotypes are determined without any post-PCR processing because probes are included in the PCR (Hui, DelMonte and Ranade 2008). Even larger-scale SNP genotyping has been conducted in *P. falciparum* using a custom 384-SNP Illumina GoldenGate with the advantage of typing numerous SNPs simultaneously. Therefore, this technology provides a practical, rapid, robust, and inexpensive assay of genome-wide parasite genotyping with easy data interpretation. The GoldenGate platform relies on

monoclonal infections to estimate allele frequencies, which is problematic when a large fraction of infections are polyclonal (Campino *et al.* 2011). The major challenge with HT-SNP genotyping is that the use of custom panels of SNPs does not allow the discovery of SNPs beyond those incorporated in the panel.

1.11.4 Whole-genome sequencing

Ever since the sequencing of the *P. falciparum* genome (Gardner *et al.* 2002), whole-genome sequencing (WGS) of field isolates has become common and has been used to study *P. falciparum* parasite diversity in natural populations (Manske *et al.* 2012). In turn, several algorithms have been developed to identify multiclonal infections and reconstruct haplotypes from sequence data. The estMOI algorithm can be used to analyze deep sequencing data for the estimation of COI, however, it is limited in that it requires deep whole genome shotgun sequence data, which is costly to generate on a large scale (Assefa *et al.* 2014).

DEploid (Zhu, Almagro-Garcia and McVean 2018) is an algorithm that analyzes genome-wide SNP data for estimation of the number of clones, their relative proportions and the haplotypes present in a sample. This method can be used to infer mixed clones with proportions at > 20% with high accuracy, but it struggles with minor clones due to insufficient read coverage. Currently, DEploid is the only software available that can reconstruct haplotypes as well as estimate COI using genome-wide SNP data. The limitation is that it requires a panel of references, which should include enough different reference clones to cover all the haplotype structure representing the field population, something that is a challenge for current computational technology.

A typical sequencing run on an Illumina NextSeq sequencer yields approximately 130 million reads of 150bp each. Even if a single isolate was sequenced by a single run, and even if coverage across the genome of 23 million bp pairs was obtained, coverage would be below 900×. Factoring in that field isolates always contain some human DNA; coverage would be even lower. Thus, when COI is the main interest of a study, whole-genome sequencing is currently not a cost-effective method.

1.11.5 Amplicon deep sequencing

Amplicon deep sequencing takes advantage of the fact that, in highly polymorphic amplicons, several SNPs may be concentrated within a locus of 100–200bp. Hence, deep sequencing of highly polymorphic makers has high sensitivity and specificity to detect minority clones in multiclonal infections. Sequencing coverage remains a challenge when estimating COI from whole-genome sequencing data. However, for amplicon deep sequencing, a read depth of 1000–10,000 results in an increased power to detect of minority clones of frequencies as low as 0.1% in experimental mixtures of different parasite clones (Lerch *et al.* 2017). The major advantage of amplicon deep sequencing over a panel of genome-wide SNPs is that all SNPs occur within one amplicon. Thus, haplotypes can be directly identified without the need for multi-locus haplotype reconstruction and software programs such as SeekDeep aid in the identification of haplotypes (Hathaway *et al.* 2018). Some of the limitations with amplicon deep-sequencing include the fact that only a small genomic region is amplified and hence is not ideal for characterizing polymorphisms that cover large genomic regions. Like point mutations, indels are another form of structural variation and reads with large indels may be aligned incorrectly or may remain unmapped (Kadri *et al.* 2015). For this reason, amplicon deep-sequencing is better suited for point mutations and shorter indels and hence may not perform well to characterize *hrp2* and *hrp3* deletions which currently relies on WGS (Sepúlveda *et al.* 2018). Also, the threshold for accurate genotype calls may be different between studies due to various sequencing error rates in different sequencing platforms. Thus, to exclude PCR or sequencing errors, it is recommended to perform experiments in duplicate and use appropriate controls so as to minimize false calls (Mideo *et al.* 2016; Hathaway *et al.* 2018).

In summary, asymptomatic infections provide a constant parasite reservoir that fuels malaria transmission and impacts on the risk of developing febrile malaria episodes. Additionally, prior studies have revealed that the COI within asymptomatic infections also impacts on the risk of developing febrile malaria. Moreover, the emergence of artemisinin and partner drug resistance in SE Asia is worrying as it may render the first line antimalarial drugs, ACTs, ineffective. For the malaria endemic SSA region, this could be catastrophic as it may lead to

increased morbidity and mortality. It is worth noting that the reduction in clinical cases appears to have stalled globally, and in some cases clinical cases have resurged, even with the deployment of effective malaria control interventions (WHO 2018a). Thus, novel interventions are required to improve the gains made in tackling malaria. Genetic epidemiology offers a resourceful tool for assessing *P. falciparum* genetic diversity to improve knowledge on the parasite diversity, its transmission dynamics, mechanisms of adaptation to environmental and interventional pressures and enables the surveillance of antimalarial drug resistance (Apinjoh *et al.* 2019).

This study, therefore, sought to examine the genetic epidemiology of the *P. falciparum* parasites in Kilifi, Kenya, using longitudinal *P. falciparum* malaria monitoring data and the genetic analysis of *P. falciparum* positive samples. This involved the examination of the epidemiology of asymptomatic infections with the aim of understanding the impact of transmission intensity and age on the risk of developing febrile malaria in asymptomatic individuals. Secondly, a genetic epidemiology analysis was conducted using amplicon deep sequencing to characterise parasite diversity in asymptomatic infections at both the individual and population level and to understand how this diversity changes during the shift to a febrile malaria episode and how it impacts of developing subsequent febrile episodes. Finally, while resistance to ACTs is yet to be detected in SSA, it is imperative that continued surveillance is carried out and molecular markers of resistance offer a resourceful tool to achieve this. Accordingly, an interrogation of 12 drug resistance markers over two decades of changing antimalarial drug policy in Kenya, was conducted. These included the validated marker for artemisinin resistance (*k13*) identified in SE Asia as well as novel markers (*ap2-mu, falcipain-2a and ubp-1*) whose impact are not yet fully understood, artemisinin resistance predisposing mutations in four genes (*arps10, crt, fd* and *mdr2*), validated markers of CQ (crt) and SP (*dhps* and *dhfr*) resistance and finally a marker (nfs) of lumefantrine resistance identified in The Gambia. The characterisation of such markers could help identify an ACT resistance genotype predictive of reduced artemisinin susceptibility in SSA, given that some studies have reported slow clearing parasites that lack *k13* mutations.

1.12 Objectives

1.12.1 Main Objective

Genetic analysis of asymptomatic *P. falciparum* infections in Kilifi, Kenya, and epidemiological evaluation of the contribution of antimalarial drug-resistance markers under changing drug selection.

1.12.2 Specific Objectives

1. To evaluate of the risk of developing febrile malaria in *P. falciparum* asymptomatic infected compared to uninfected children in Kilifi.
2. To use deep-sequencing to evaluate *P. falciparum* diversity in asymptomatic and febrile infections in Kilifi.
3. To assess the genetic diversity of 12 *P. falciparum* drug-resistance markers (*crt*, *mdr1*, *dhps*, *nfs*, *k13*, *ap2mu*, *falcipain-2a*, *ubp-1*, as well as four artemisinin resistance predisposing mutations; *arps10* codon V127M*, crt* codon I356T*, fd* codon D193Y and *mdr2* codon T484I), over two decades of changing antimalarial drug policy in Kilifi, Kenya.

1.13 Justification

P. falciparum asymptomatic infections may negatively impact malaria elimination strategies since they become a constant source of parasites that fuel malaria transmission. Consequently, it has been suggested that in addition to targeting febrile infections, asymptomatic infections should also be targeted to have a significant effect on malaria transmission. However, there is conflicting literature on what the role of asymptomatic infections in on the risk of developing febrile episodes. Also, if asymptomatic infections are associated with a reduced risk of febrile episodes, what will the impact of clearing these infections be on the risk of developing febrile episodes at a population level. For this reason, this study sought to evaluate how asymptomatic infections impact on the risk of developing febrile malaria in Kilifi, as it will inform malaria intervention strategies.

High genetic diversity has been reported to correlate with malaria transmission intensity (Kateera *et al.* 2016; Nabet *et al.* 2016). Like many places in Africa, Kilifi has seen a drop in

malaria transmission intensity from the 1990s to 2008, however, it has since started to increase again (Snow *et al.* 2015). What effect this drop and recent increase in transmission intensity has had on the genetic diversity of the local parasite population, has not been evaluated. This study sought to use amplicon deep-sequencing to assess parasite diversity over time, in parallel with changing malaria transmission intensities, in both asymptomatic and febrile infections.

Finally, reports of emergence and spread of artemisinin resistance in SE Asia necessitates continued surveillance of drug-resistance markers in regions yet to report artemisinin resistance, such as SSA. Studies conducted in SSA to date have not detected the *k13* mutations that have been associated with artemisinin resistance, but some have reported mutations in other markers such as *ap2-mu* and *falcipain-2a* that need to be evaluated further. This study, therefore, sought to evaluate the prevalence of *k13*, *ap2-mu* and *falcipain-2a* mutations in Kilifi and assess how these have changed over two decades of changing antimalarial policy.

1.14 Ethical approval

Ethical approval for these studies was obtained from the Ethics Review Committee of the Kenya Medical Research Institute under protocol number SERU 3149.

Chapter 2 *: P. falciparum* Asymptomatic Infections and the Risk of Developing Febrile Malaria

2.1 Introduction

As previously highlighted in chapter 1, asymptomatic infections are detrimental to the host, with serious health, developmental, and productivity consequences. Additionally, asymptomatic infections can later transition to febrile malaria and the factors that influence this are yet to be fully understood. Using microscopy for diagnosis, some studies have associated asymptomatic infections with a lesser risk of developing febrile malaria. This apparent protection has been linked to the fact that individuals living in high malaria transmission settings have higher levels of immunity compared to those in low transmission settings (Doolan, Dobaño and Baird 2009). As highlighted in section 1.9.3, there is conflicting literature on the role of asymptomatic infections on the risk of developing febrile malaria. In a study of 316 children aged 6 months to 5 years in an area of moderate malaria transmission in Uganda, children with asymptomatic infections were at a higher risk of febrile malaria compared to uninfected children within the first 30 days of detection (Njama-Meya, Kamya and Dorsey 2004). Likewise, in a study of 566 children aged 2 to 17 years in an area of low malaria transmission in Senegal, it was shown that asymptomatic infections at the beginning of the transmission season were associated with increased febrile malaria episodes, independently of age. However, this association was not observed at the end of the transmission season (Le Port *et al.* 2008). In contrast, in a study of 610 children under 6 years of age in an area of high malaria transmission in Tanzania, younger children (<3 years of age) with asymptomatic infections at baseline were observed to be at a higher risk of febrile malaria while older children (>3 years of age) appeared to be at a reduced risk (Henning *et al.* 2004). Similarly, in a larger study of 1356 children aged 2 to 18 years in an area of high malaria transmission in Senegal, it was shown that children with asymptomatic infections in the dry season leading into the rainy season were at a lower risk of developing febrile malaria. Additionally, this association was age-dependent as younger children (≤5 years of age) were found to be at a higher risk than older children (>5 years of age) (Males, Gaye and Garcia 2008). In a study of 695 individuals aged 6 months to 25 years in an area of high malaria transmission in Mali, participants with asymptomatic infections at the end of the dry season were at a lower risk of febrile malaria. The study also found that older individuals experienced

less febrile malaria than younger children, although no age stratification data was provided (Portugal *et al.* 2017). Finally, in a study of 120 participants aged between 1-50 years in an area of high malaria transmission in Malawi, participants were treated with AL at enrolment and followed for 2 years with monthly routine visits. It was shown that asymptomatic infections were associated with decreased risk of malaria illness across all ages (Buchwald *et al.* 2018).

The above-mentioned studies have shown that in addition to asymptomatic infections having an impact on the risk of developing subsequent febrile episodes, malaria transmission intensity also appears to have a role. In line with this, it has been shown that immunity to malaria is acquired more rapidly with increasing transmission. Therefore, individuals harbouring asymptomatic infections may be protected from malaria because of increased exposure. Regarding the role of age and protection from malaria, there may be something intrinsic about age other than the constant exposure for example a maturing immunity. To shed light on this, one has to explore the limitations of previous studies that examined the impact of asymptomatic infections and risk of malaria. Notably, these studies were conducted in different geographical regions and with participants of varying age groups. Therefore, conclusions based on the assumption that differences in transmission intensity and age explain the opposing results, require confirmation. In this chapter, the impact of both malaria transmission intensity and age on the risk of developing febrile malaria in children with asymptomatic infections, was evaluated. These findings may provide insights into the impact of immunity on the risk of developing febrile malaria in asymptomatic individuals living in different transmission settings.

2.2 Methods

2.2.1 Study design and data collection

The data used for this analysis is based on three cohorts in Kilifi of varying malaria transmission intensities, including Ngerenya (low transmission), Junju (moderate to high transmission) and Chonyi (high transmission). The data were prospectively collected between 1998 and 2014 for Ngerenya, 2005 and 2016 for Junju, and 1999 and 2001 for Chonyi. In these

cohorts, children were recruited at birth for weekly clinical malaria monitoring until the age of 15 years (Mwangi *et al.* 2005; Ndungu *et al.* 2015).

The transmission intensity in Kilifi was on a general decline during the study period, from 1989 to 2010 and later started to increase (Snow *et al.* 2015; Mogeni *et al.* 2016), however, it was much higher in Junju and Chonyi than in Ngerenya. Data from the active weekly surveillance was used to determine malaria episodes and annual cross-sectional surveys to define asymptomatic infections and uninfected children. There are two rainy seasons per year in Kenya during which malaria transmission increases, the long rains that occur in May-July and the short rains that occur in October-November (Mwangi *et al.* 2005). The cross-sectional surveys are conducted before the long rains.

2.2.2 Case definitions

Asymptomatic infection was defined as having any detectable *P. falciparum* parasites by microscopy (thick and thin smears), having an axillary temperature <37.5°C and no history of fever during the cross-sectional survey. Additionally, the children should not have had a recent febrile malaria episode within the month before the survey or had a fever within the subsequent 7 days from the date of the survey. Uninfected children were defined as those without *P. falciparum* parasitemia and no fever during the cross-sectional survey. A febrile malaria episode was defined as having ≥2500 parasites/μl and an axillary temperature greater than 37.5°C, based on parasite cut-offs previously defined for these cohorts (Mwangi *et al.* 2005).

2.2.3 Statistical analysis

Survival analysis was used to compare the risk of developing febrile malaria between the asymptomatic and uninfected children and survival times were estimated as follows: the baseline (time 0) was defined as the date of cross-sectional survey while the end of follow up was set at the last visit or at 365 days if follow up continued to the subsequent annual cross-sectional survey; the time to first febrile malaria episode was defined as the interval between the baseline date and the first febrile malaria episode. Children who did not complete follow up or did not get a febrile episode by the end of follow up were censored.

Kaplan-Meier curves and log-rank test were used to compare time to first febrile malaria episode. The Cox proportional hazards model was used to investigate the association between the risk of developing a febrile episode and the following covariates: transmission intensity and asymptomatic infections as categorical variables, age as a continuous linear and non-linear variable (after transformation using multiple fractional polynomials). The calendar year was adjusted for as a continuous variable to account for falling transmission intensity and the non-independence of repeated observations was accounted for using the robust cluster Huber-White method. The proportional hazards assumption was tested for using Schoenfeld residuals and finally, the model included interactions with time for the covariates that showed non-proportional hazards. Model fits were assessed using log-likelihood, Akaike's information criterion (AIC) and Bayesian information criterion (BIC) scores and the model with the least scores was chosen to be the one that best fit the data.

Data cleaning, plotting of Kaplan-Meier curves and log-rank tests were carried out in R v3.5.1 using the following R packages: data.table v1.11.4 (Dowle and Srinivasan 2018), dplyr v0.7.6 (Wickham *et al.* 2018), survival v2.42-6 (Therneau 2020) and survminer v0.4.3 (Kassambara and Kosinski 2018). Cox regression analysis was performed in Stata v14.0.

2.3 Results

2.3.1 Demography

A total of 2,644 children were recruited into the study over 17 cross-sectional surveys. 20 children were excluded due to missing data on age, bringing the person-years of follow up to 12,543. 1,218 person-years of follow up were excluded from the cross-sectional surveys for the following reasons; 249 lacked data on parasitemia or temperature, 220 had a febrile non-malaria episode, 99 had febrile malaria, 139 asymptomatics who did not meet the criteria for defining an asymptomatic infection (see section 2.2.2 regarding case definitions) and 511 lacked weekly surveillance data. Consequently, data on 2,513 children were analysed, representing 11,325 person-years of follow up (**Figure 2.1**). There was a total of 3,132 episodes of malaria during the cross-sectional surveys, while 15.2% of children had asymptomatic infections across all sites (**Table 2.1**). Finally, **Table 2.2** shows the distribution of the asymptomatic and uninfected episodes in the three cohorts from 1999 to 2016.

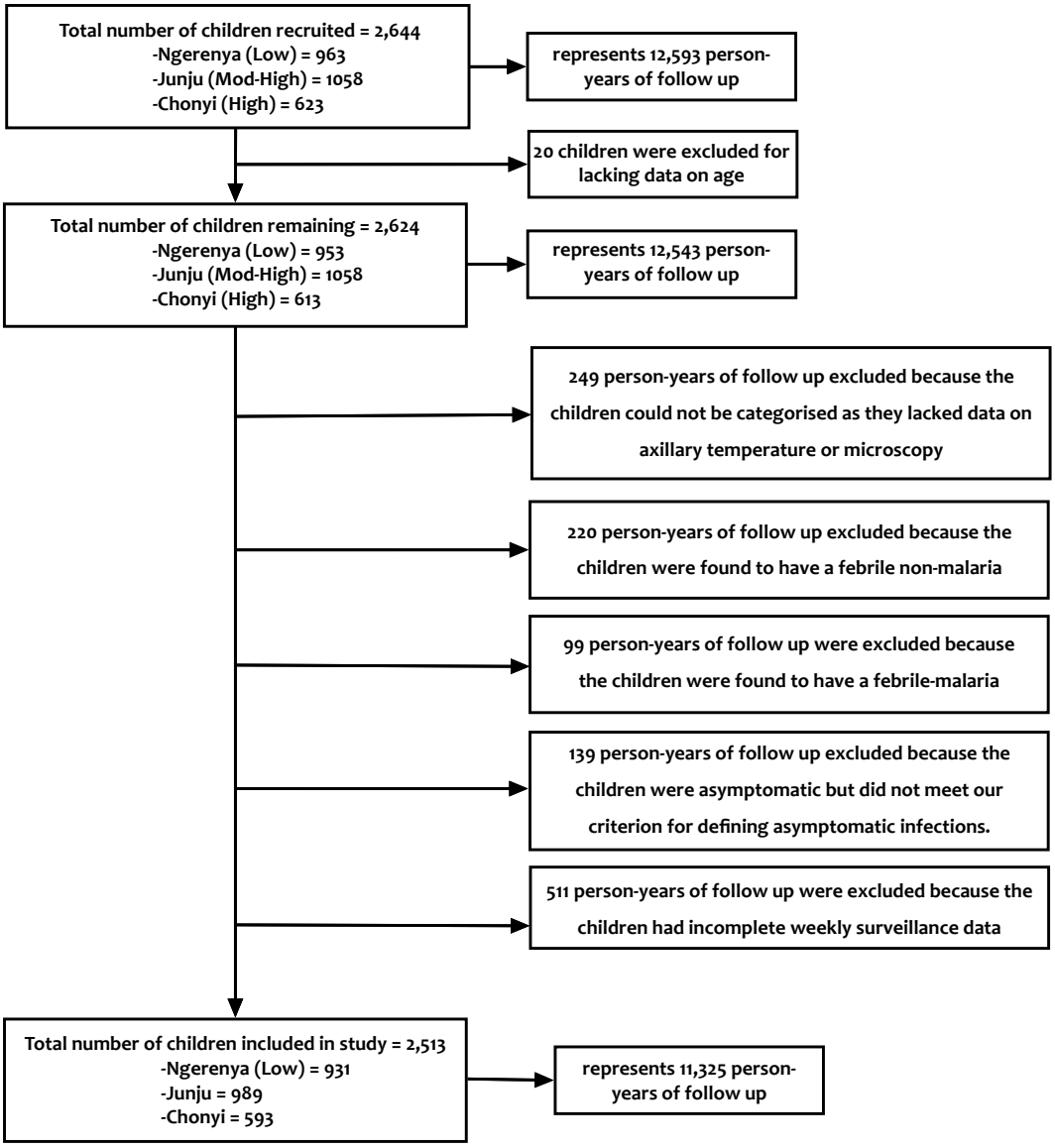


Figure 2.1 A flow chart showing the total number of children recruited from three cohorts of varying malaria transmission intensity (Ngerenya – low, Junju – moderate to high and Chonyi – high) and the numbers of children that met the case definitions for this study.

Table 2.1. The Demography of the study participants from the three cohorts

	Cohort		
	Ngerenya (Low)	**Junju (Mod-High)**	**Chonyi (High)**
Total Number of children included in the study	931	989	593
Age range [Years]	>0 - 15	>0 – 15	>0 - 15
Person years of follow-up	5,179	5,819	1,595
Total Number of males [%]	484 [51.99]	497 [50.25]	306 [51.60]
Total Number asymptomatic episodes [%]	419 [9.50%]	923 [17.00%]	538 [36.40%]
Total number of children with 1 or more febrile malaria episodes [%]	162 [17.40]	371 [37.51]	119 [20.06]

Low – low transmission, *Mod-High* – moderate to high transmission, *High* – high transmission

Table 2.2. Distribution of asymptomatic and uninfected cases through the years in the three cohorts

Year	**Ngerenya (Low)**		**Junju (Mod-High)**		**Chonyi (High)**	
	Asymptomatic	**Uninfected**	**Asymptomatic**	**Uninfected**	**Asymptomatic**	**Uninfected**
1999	128	360	-	-	233	218
2000	73	453	-	-	143	372
2001	125	405	-	-	162	350
2002	35	247	-	-	-	-
2003	33	237	-	-	-	-
2004	15	261	-	-	-	-
2005	6	245	167	-	-	-
2006	-	-	-	-	-	-
2007	2	255	50	275	-	-
2008	1	270	85	242	-	-
2009	0	240	84	570	-	-
2010	0	257	128	523	-	-
2011	0	261	96	562	-	-
2012	1	259	94	574	-	-
2013	0	241	48	623	-	-
2014	-	-	79	568	-	-
2015	-	-	59	306	-	-
2016	-	-	33	271	-	-
Totals	***419***	***3991***	***923***	***4514***	***538***	***940***

The time-points range from 1999 through to 2016, however, in 2006 no survey was conducted as the malaria monitoring study transitioned to a new protocol that began in 2007. Marked with "-" are the 2014-2016 time points for Ngerenya when no survey was conducted since malaria transmission was very low and surveillance switched from active to passive. Sample collection stopped in Chonyi after 2001 and the new study site was Junju. Additionally, for the Junju 2005 time-point, the survey included asymptomatic children only. *Low* – low transmission, *Mod-High* – moderate to high transmission, *High* – high transmission.

2.3.2 Risk of developing febrile malaria

The risk of developing febrile malaria was compared between children with asymptomatic infections and uninfected children at baseline. Uninfected children were found to be at a higher risk compared to children with asymptomatic infections from the start of follow up and this difference in rates of developing episodes was more distinct after 90 days, when the two survival functions diverged to the end of follow up (**Figure 2.2**, p-value < 0.0001).

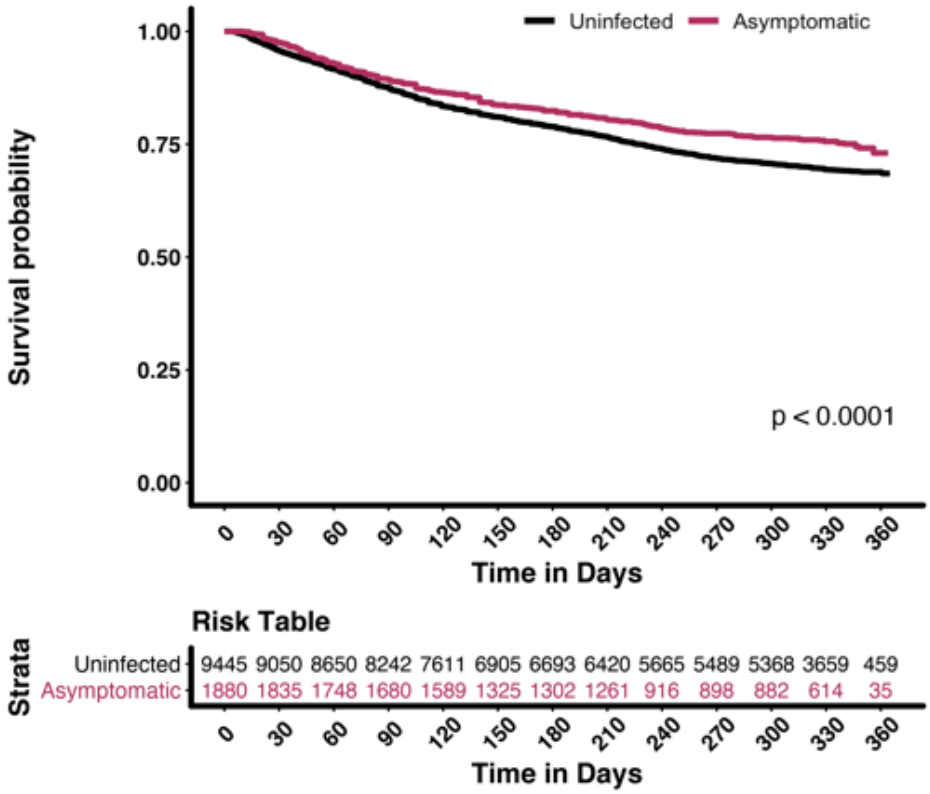


Figure 2.2. A Kaplan-Meier plot showing the risk of developing febrile malaria in uninfected children vs. children with asymptomatic infections across three malaria transmission settings. This plot compares the time to first febrile malaria episode between uninfected children versus children with asymptomatic infections across the three cohorts of varying malaria transmission intensities over a one-year period. The x-axis represents the one year longitudinal follow up from cross-sectional survey until one developed a febrile episode while the y-axis represents the proportion of individuals that have not yet developed a febrile episode. The risk table shows the number of participants under observation at every 30-days interval across the two groups. The log-rank test was used to compare the survival distributions between the two groups.

The effect of age and transmission intensity on the risk of developing febrile malaria in children with asymptomatic infections vs. uninfected children was then examined. Asymptomatic infections were associated with an increased risk of febrile malaria within all age groups in the low transmission setting. The risk of febrile malaria was not altered by asymptomatic infections among the younger age groups (≤ 3 years) in the moderate to high and in the high transmission settings. However, among the older age groups (> 3 years) in the moderate-high and high transmission settings, asymptomatic infections were associated with a reduced risk of developing febrile malaria (**Figure 2.3**).

There were some variations in effect over time including, the risk among children with asymptomatic infections was found to decrease over time relative to the risk among children without asymptomatic infection. In turn, the divergence between the survival functions occurred in earlier time periods in settings where asymptomatic infections were associated with increased risk (i.e. in children of all ages in the lower transmission region) and became parallel in later time periods. Where asymptomatic infections were associated with reduced risk (i.e. in older children in the higher transmission region) the survival functions tended to be parallel in earlier time periods (or even showed increased risk among children with asymptomatic infection) and then showed decreased risk among children with asymptomatic infection during later time periods (**Figure 2.3**).

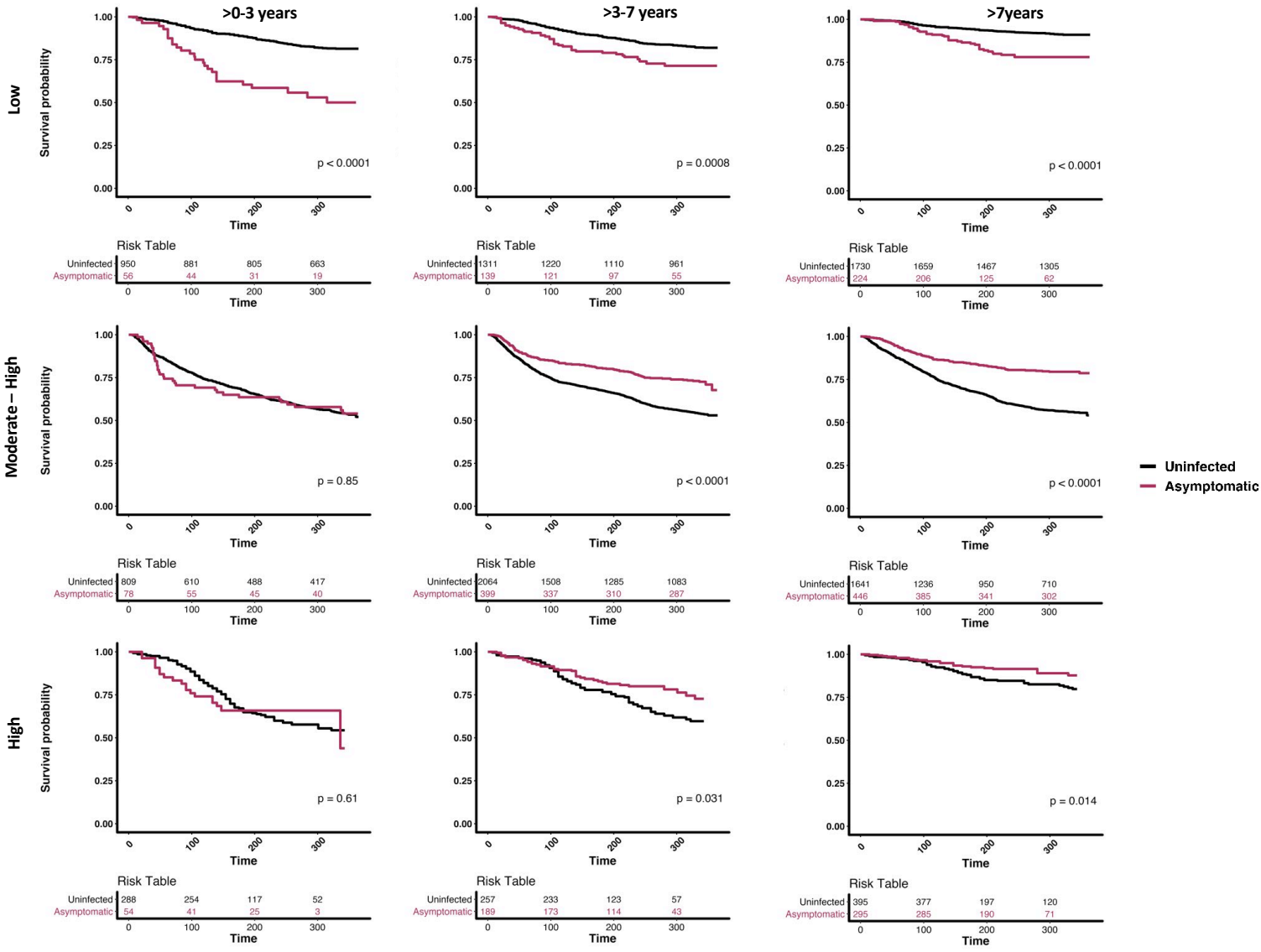
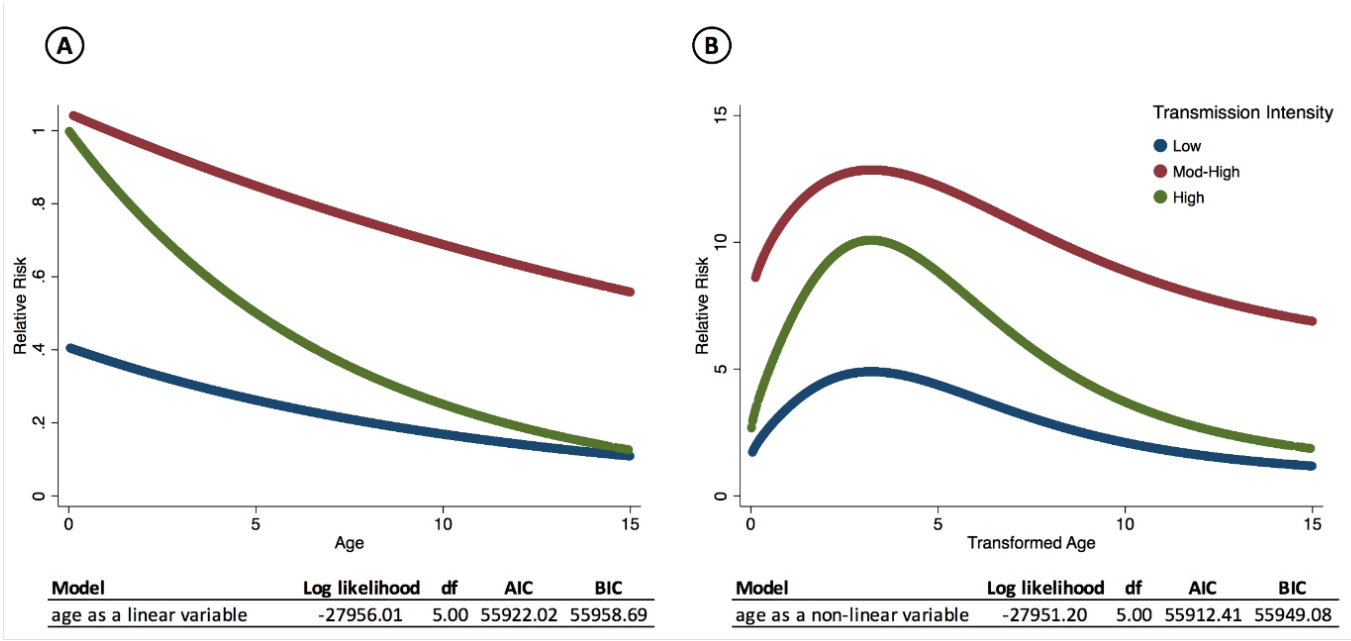


Figure 2.3. Risk of developing febrile malaria between asymptomatic and uninfected children, stratified by malaria transmission intensity and age. This plot compares the time to the first febrile malaria episode between uninfected children versus children with asymptomatic infections across the three cohorts of varying malaria transmission intensities, stratified by three age groups, 0-3, >3-7 and >7 years old. The risk table shows the number of participants under observation at 100-day intervals for both the uninfected (black) and asymptomatic (maroon) groups. The log-rank test was used to compare survival distributions.

2.3.3 The impact of transmission intensity and age on the risk of developing febrile malaria

The multiple fractional polynomials (MFP) approach was used to transform age as a non-linear variable and this was found to best fit the data compared to having age as a linear

variable ($p < 0.0001$, **Figure 2.4**). MFP are used for modelling the influence of continuous variables on the outcome of interest in regression models (Royston and Altman 1994; Sauerbrei and Royston 1999). These models are useful when one wishes to preserve the continuous nature of the covariates in a regression model, but suspects that some or all of the relationships may be non-linear. In this case, it was suspected that the relationship between age and risk of subsequent malaria episodes may not follow a linear pattern, hence further models were constructed with the transformed age variable.



Model	Log likelihood	df	AIC	BIC
age as a linear variable	-27956.01	5.00	55922.02	55958.69

Model	Log likelihood	df	AIC	BIC
age as a non-linear variable	-27951.20	5.00	55912.41	55949.08

Figure 2.4 Models comparing the relative risk of developing febrile episodes as predicted from the interaction between site and age as (A) linear and (B) non-linear (transformed using fractional polynomials). The log-likelihood, AIC and BIC scores below each respective graph show that the non-linear model is a better fit since it had the least scores. For this reason, further models were constructed with non-linear age.

After incorporating age as a non-linear variable, a Cox proportional hazards model was constructed to test the effect of transmission intensity, infection status (Asymptomatic vs. Uninfected), transformed age, gender (male vs. female), year of survey and time cut-offs (days where the two survival functions diverge) and their interactions. Moreover, an examination was conducted to compare several multivariable models that allowed for linear

and logarithmic interactions between time and other covariates (transmission, transformed age, infection status and year of survey).

The first model tested the effect of all the covariates on the risk of developing febrile malaria and time was included as exponential. The following covariates were associated with an increased risk including transmission intensity (mod-high vs. high, hazard ratio [HR] = 15.92, p-value < 0.0001), transformed age (HR = 5.06, p-value < 0.0001), the interaction between transmission intensity (low vs. high) and asymptomatic infections (HR = 2.38, p-value < 0.0001), the interaction between asymptomatic infections and transformed age (HR = 3.44, p-value <0.0001), the interaction between asymptomatic infections and year of survey (HR = 1.13, p-value <0.0001) and gender though not significant (HR = 1.06, p-value = 0.246). On the other hand, the following covariates were associated with a reduced risk including transmission intensity (low vs. high) though not significant (HR = 0.92, p-value = 0.815), the interaction between transmission intensity (low vs. high) and transformed age though not significant (HR = 0.65, p-value = 0.213), the interaction between transmission intensity (mod-high vs. high) and transformed age (HR = 0.25, p-value < 0.0001), asymptomatic infections (HR = 0.16, p-value < 0.0001), the interaction between transmission (mod-high vs. high) and asymptomatic infections (HR = 0.24, p-value < 0.0001) and finally the year of survey (HR = 0.92, p-value = <0.0001) (**Table 2.3**).

Table 2.3. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria, without time varying covariates

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Transmission (Low vs. High)	0.92	0.32	-0.23	0.815	0.47	1.82
Transmission (Mod-High vs. High)	15.92	5.16	8.54	**<0.0001**	8.44	30.06
Transformed Age	5.06	1.40	5.86	**<0.0001**	2.94	8.71
Transmission (Low vs. High) x Transformed Age	0.65	0.23	-1.24	0.213	0.33	1.28
Transmission (Mod-High vs. High) x Transformed Age	0.25	0.08	-4.58	**<0.0001**	0.13	0.45
Infection Status (Asymptomatic vs. Uninfected)	0.16	0.05	-5.75	**<0.0001**	0.09	0.30
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.38	0.44	4.74	**<0.0001**	1.66	3.41
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.24	0.06	-5.46	**<0.0001**	0.15	0.40
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.44	0.98	4.35	**<0.0001**	1.97	6.02
Gender (Male vs. Female)	1.06	0.05	1.16	0.246	0.96	1.16
Year of Survey	0.92	0.01	-10.25	**<0.0001**	0.90	0.93
Infection Status (Asymptomatic vs. Uninfected) x Year of Survey	1.13	0.02	5.6	**<0.0001**	1.08	1.18

The odds ratio at baseline for all the covariates are presented. Time was included as exponential. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

From the Schoenfeld's residuals test, the relative effect of infection status and year of survey on the hazard function was found to change over time. Thus, this violated the proportional hazards assumption which requires that a covariate effect is constant over time (**Table 2.4**) Further models, therefore, included infection status and year of survey as time-varying covariates.

Table 2.4. Schoenfeld's residuals for the tested covariates

Co-variate	rho	chi2	df	Prob>chi2
Infection Status (Asymptomatic vs. Uninfected)	-0.041	7.26	1	**0.007**
Transmission (Low vs. High)	-2×10^{4}	<0.01	1	0.99
Transmission (Mod-High vs. High)	-0.024	2.54	1	0.11
Age	-0.003	0.04	1	0.85
Gender (Male vs. Female)	-0.008	0.42	1	0.52
Year of Survey	-0.058	18.26	1	**<0.0001**
Global Test		95.61	6	**<0.0001**

The global p-value shows that the effect of at least one of the covariates varies over time. Looking at each of the covariates independently, the effects of infection status and year of survey vary over time as they had p-values that were below 0.05. *Abbreviations: mod-high – moderate-high.*

The second model tested the effect of all covariates on the risk of developing febrile malaria and included time-varying covariates as well as time as an exponential variable. Similar effects of the covariates were observed as in the first model. Also, among the time varying covariates, both asymptomatic infections and year of survey had no impact on risk (HR = 1) (**Table 2.5**).

Table 2.5. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria, with time varying covariates and time as an exponential variable

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	0.99	0.35	-0.04	0.972	0.49	1.98
Transmission (Mod-High vs. High)	16.69	5.49	8.56	**<0.0001**	8.76	31.80
Transformed Age	5.37	1.53	5.90	**<0.0001**	3.07	9.38
Transmission (Low vs. High) x Transformed Age	0.59	0.21	-1.51	0.132	0.29	1.18
Transmission (Mod-High vs. High) x Transformed Age	0.23	0.07	-4.67	**<0.0001**	0.13	0.43

Infection Status (Asymptomatic vs. Uninfected)	0.21	0.07	-4.96	**<0.0001**	0.11	0.39
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.47	0.46	4.89	**<0.0001**	1.72	3.56
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.26	0.07	-5.23	**<0.0001**	0.15	0.43
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.33	0.94	4.25	**<0.0001**	1.91	5.78
Gender (Male vs. Female)	1.06	0.05	1.17	0.241	0.96	1.16
Year of Survey	0.95	0.01	-4.74	**<0.0001**	0.94	0.97
Infection Status (Asymptomatic vs. Uninfected) x Year of Survey	1.12	0.02	5.46	**<0.0001**	1.08	1.17
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	1.00	0.00	-2.75	**0.006**	1.00	1.00
Year	1.00	0.00	-6.83	**<0.0001**	1.00	1.00

The odds ratio at baseline (main) for all the covariates and those that varied over time (time-varying covariates) are presented. Time was included as exponential. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

The third model was similar to the second model; however, gender was dropped as it was found not to have a significant effect in the first and second models. Similar effects of the other covariates were observed as in the second model (**Table 2.6**).

Table 2.6. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria after dropping gender from the model, with time varying covariates and time as an exponential variable

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	0.99	0.35	-0.03	0.977	0.49	1.98
Transmission (Mod-High vs. High)	16.58	5.46	8.52	**<0.0001**	8.69	31.62
Transformed Age	5.36	1.53	5.89	**<0.0001**	3.07	9.38
Transmission (Low vs. High) x Transformed Age	0.59	0.21	-1.50	0.133	0.29	1.18
Transmission (Mod-High vs. High) x Transformed Age	0.23	0.07	-4.64	**<0.0001**	0.13	0.43
Infection Status (Asymptomatic vs. Uninfected)	0.21	0.07	-4.93	**<0.0001**	0.11	0.39
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.46	0.46	4.86	**<0.0001**	1.71	3.55
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.26	0.07	-5.25	**<0.0001**	0.15	0.42
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.31	0.93	4.24	**<0.0001**	1.90	5.75
Year	0.95	0.01	-4.73	**<0.0001**	0.94	0.97
Infection Status (Asymptomatic vs. Uninfected) x Year	1.12	0.02	5.47	**<0.0001**	1.08	1.17
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	1.00	0.00	-2.74	**0.006**	1.00	1.00
Year of Survey	1.00	0.00	-6.84	**<0.0001**	1.00	1.00

The odds ratio at baseline (main) for all the covariates and those that varied over time (time-varying covariates) are presented. Also, gender was dropped as it was found not to be significant and time was included as exponential. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

The fourth model was similar to the third model but time was log-transformed, to compare with time as an exponential variable and the effects of the covariates were similar to the third model. However, among the time varying covariates, the effects of asymptomatic infections

and year of survey were associated with a non-significant and significant reduced risk, respectively (HR = 0.81, p-value = 0.09 and HR = 0.93, p-value < 0.0001, respectively) (**Table 2.7**).

Table 2.7. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria, with time varying covariates and time as a log-transformed variable

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	0.98	0.34	-0.07	0.945	0.49	1.95
Transmission (Mod-High vs. High)	16.44	5.39	8.53	**<0.0001**	8.64	31.28
Transformed Age	5.27	1.49	5.87	**<0.0001**	3.03	9.18
Transmission (Low vs. High) x Transformed Age	0.60	0.21	-1.43	0.154	0.30	1.21
Transmission (Mod-High vs. High) x Transformed Age	0.24	0.07	-4.60	**<0.0001**	0.13	0.44
Infection Status (Asymptomatic vs. Uninfected)	0.26	0.10	-3.52	**<0.0001**	0.12	0.55
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.42	0.45	4.76	**<0.0001**	1.68	3.48
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.25	0.06	-5.36	**<0.0001**	0.15	0.41
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.33	0.94	4.26	**<0.0001**	1.92	5.81
Year of Survey	1.05	0.02	2.05	**0.041**	1.00	1.10
Infection Status (Asymptomatic vs. Uninfected) x Year of Survey	1.13	0.02	5.56	**<0.0001**	1.08	1.17
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	0.81	0.10	-1.69	0.09	0.63	1.03
Year of Survey	0.93	0.01	-5.90	**<0.0001**	0.91	0.95

The odds ratio at baseline (main) for all the covariates and those that varied over time (time-varying covariates) are presented. Also, gender was dropped as it was found not to be significant and time was included as log-transformed. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

Time as an exponential variable was found to be a better fit for the data, compared with time as a log-transformed variable. Therefore, subsequent models maintained time as an exponential variable. Additionally, the Kaplan-Meier plots revealed that there were variations in effect on risk over time. As mentioned earlier, this resulted in the divergence between the survival functions occurring in earlier time periods in settings where asymptomatic infections were associated with increased risk and became parallel in later time periods. Additionally, where asymptomatic infections were associated with reduced risk, the survival functions tended to be parallel in earlier time periods and then showed decreased risk among children with asymptomatic infection during later time periods. Consequently, the fifth (**Table 2.8**), sixth (**Table 2.9**), seventh (**Table 2.10**) and eighth models (**Table 2.11**) tested the effect of all covariates on the risk of developing febrile malaria and included step-wise interactions between time (at cut-offs of 30, 60, 90, and 120 days, respectively) and other covariates.

Similar effects of the covariates were observed as in the fourth model. The comparison of goodness-of-fit statistics is presented in **Table 2.12**.

Table 2.8. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria, with time varying covariates and time as an exponential variable with alternative cut-offs of <30 and >30 days

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	0.94	0.33	-0.19	0.849	0.47	1.85
Transmission (Mod-High vs. High)	15.95	5.19	8.51	**<0.0001**	8.43	30.18
Transformed Age	5.09	1.42	5.84	**<0.0001**	2.95	8.78
Transmission (Low vs. High) x Transformed Age	0.64	0.22	-1.27	0.204	0.32	1.27
Transmission (Mod-High vs. High) x Transformed Age	0.25	0.08	-4.56	**<0.0001**	0.13	0.45
Infection Status (Asymptomatic vs. Uninfected)	0.13	0.05	-5.83	**<0.0001**	0.07	0.26
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.37	0.44	4.67	**<0.0001**	1.65	3.40
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.24	0.06	-5.50	**<0.0001**	0.14	0.40
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.44	0.98	4.33	**<0.0001**	1.97	6.01
Year of Survey	0.94	0.01	-5.15	**<0.0001**	0.92	0.96
Infection Status (Asymptomatic vs. Uninfected) x Year of Survey	1.13	0.02	5.68	**<0.0001**	1.08	1.18
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	1.26	0.21	1.39	0.164	0.91	1.73
Year of Survey	0.97	0.01	-2.80	**0.005**	0.95	0.99

The odds ratio at baseline (main) for all the covariates and those that varied over time (time-varying covariates) are presented. Also, gender was dropped as it was found not to be significant, time was included as exponential with alternative cut-offs of <30 and >30 days. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

Table 2.9. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria, with time varying covariates and time as an exponential variable with alternative cut-offs of <60 and >60 days

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	0.96	0.34	-0.11	0.912	0.48	1.92
Transmission (Mod-High vs. High)	16.37	5.35	8.55	**<0.0001**	8.63	31.07
Transformed Age	5.20	1.46	5.86	**<0.0001**	3.00	9.03
Transmission (Low vs. High) x Transformed Age	0.62	0.22	-1.35	0.176	0.31	1.24
Transmission (Mod-High vs. High) x Transformed Age	0.24	0.08	-4.57	**<0.0001**	0.13	0.44
Infection Status (Asymptomatic vs. Uninfected)	0.22	0.07	-4.82	**<0.0001**	0.12	0.40
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.40	0.44	4.76	**<0.0001**	1.67	3.45
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.25	0.06	-5.38	**<0.0001**	0.15	0.41
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.35	0.94	4.29	**<0.0001**	1.93	5.82
Year of Survey	0.96	0.01	-4.77	**<0.0001**	0.94	0.97
Infection Status (Asymptomatic vs. Uninfected) x Year of Survey	1.12	0.02	5.43	**<0.0001**	1.08	1.17
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	0.72	0.08	-2.92	**0.003**	0.57	0.90
Year of Survey	0.94	0.01	-6.50	**<0.0001**	0.92	0.96

The odds ratio at baseline (main) for all the covariates and those that varied over time (time-varying covariates) are presented. Also, gender was dropped as it was found not to be significant, time was included as exponential with alternative cut-offs of <60 and >60 days. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

Table 2.10. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria, with time varying covariates and time as an exponential variable with alternative cut-offs of <90 and >90 days

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	0.98	0.35	-0.05	0.957	0.49	1.96
Transmission (Mod-High vs. High)	16.55	5.43	8.55	**<0.0001**	8.70	31.49
Transformed Age	5.29	1.50	5.86	**<0.0001**	3.03	9.22
Transmission (Low vs. High) x Transformed Age	0.60	0.21	-1.43	0.154	0.30	1.21
Transmission (Mod-High vs. High) x Transformed Age	0.24	0.07	-4.59	**<0.0001**	0.13	0.44
Infection Status (Asymptomatic vs. Uninfected)	0.21	0.07	-4.97	**<0.0001**	0.12	0.39
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.43	0.45	4.81	**<0.0001**	1.69	3.49
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.25	0.07	-5.29	**<0.0001**	0.15	0.42
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.32	0.93	4.27	**<0.0001**	1.91	5.76
Year of Survey	0.95	0.01	-5.72	**<0.0001**	0.93	0.97
Infection Status (Asymptomatic vs. Uninfected) x Year of Survey	1.12	0.02	5.35	**<0.0001**	1.07	1.17
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	0.69	0.08	-3.39	**0.001**	0.56	0.85
Year of Survey	0.94	0.01	-7.34	**<0.0001**	0.92	0.95

The odds ratio at baseline (main) for tall the covariates and those that varied over time (time-varying covariates) are presented. Also, gender was dropped as it was found not to be significant, time was included as exponential with alternative cut-offs of <90 and >90 days. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

Table 2.11. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria, with time varying covariates and time as an exponential variable with alternative cut-offs of <120 and >120 days

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	0.96	0.34	-0.11	0.909	0.48	1.92
Transmission (Mod-High vs. High)	16.18	5.30	8.49	**<0.0001**	8.51	30.76
Transformed Age	5.24	1.48	5.86	**<0.0001**	3.01	9.12
Transmission (Low vs. High) x Transformed Age	0.61	0.22	-1.40	0.161	0.30	1.22
Transmission (Mod-High vs. High) x Transformed Age	0.24	0.07	-4.59	**<0.0001**	0.13	0.44
Infection Status (Asymptomatic vs. Uninfected)	0.18	0.06	-5.41	**<0.0001**	0.10	0.34
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.44	0.45	4.82	**<0.0001**	1.70	3.50
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.25	0.07	-5.29	**<0.0001**	0.15	0.42
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.35	0.95	4.27	**<0.0001**	1.92	5.82
Year of Survey	0.93	0.01	-7.84	**<0.0001**	0.92	0.95
Infection Status (Asymptomatic vs. Uninfected) x Year of Survey	1.12	0.02	5.52	**<0.0001**	1.08	1.17
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	0.79	0.09	-2.03	**0.042**	0.64	0.99
Year of Survey	0.95	0.01	-5.06	**<0.0001**	0.94	0.97

The odds ratio at baseline (main) for all the covariates and those that varied over time (time-varying covariates) are presented. Also, gender was dropped as it was found not to be significant, time was included as exponential with alternative cut-offs of <120 and >120 days. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. *Abbreviations: mod-high – moderate-high.*

In the end, the model with alternative cut-offs of <90 and >90 days (**Table 2.10**) was found to best fit the data as it had the least log-likelihood, AIC and BIC scores (**Table 2.12**) and was retained as the final model.

Table 2.12. Markers of goodness-of-fit for Cox proportional hazards models

Model	Reference	Log-likelihood	df	AIC	BIC
A multivariable model with age as non-linear and its interactions, no tvc	Table 2.3	-27787.36	12	55598.72	55686.73
A multivariable model with age as non-linear and its interactions, with tvc	Table 2.5	-27756.43	14	55540.86	55643.54
A multivariable model with time-varying covariates (statistically significant covariates only)	Table 2.6	-27757.68	13	55541.36	55636.71
A multivariable model with time-varying covariates (logarithmic variations with time)	Table 2.7	-27763.78	13	55553.55	55648.90
A multivariable model with time-varying covariates (cut-offs)	30-day cut-off (Table 2.8)	-27782.10	13	55590.20	55685.55
	60-day cut-off (Table 2.9)	-27761.11	13	55548.23	55643.58
	90-day cut-off (Table 2.10)	-27750.39	13	55526.79	55622.14
	120-day cut-off (Table 2.11)	-27770.23	13	55566.47	55661.82

The log-likelihood score is a statistical test used for comparing the goodness of fit of two models. Based on the likelihood ratio, it expresses how many times more likely the data are under one model than the other and smaller log-likelihood values indicate a better model. The AIC and the BIC scores are information-based criteria that assess the model fit. In general, BIC penalizes models with more parameters more than AIC does. For this reason, it leads to choosing more parsimonious models, that is, models with fewer parameters, than does AIC. In both cases, lower AIC and BIC scores indicate a better model. *Abbreviations: df – degrees of freedom, tvc – time-varying covariates.*

In the final model, the following covariates were associated with an increased risk including transmission intensity (mod-high vs. high, hazard ratio [HR] = 16.55, p-value < 0.0001), transformed age (HR = 5.29, p-value < 0.0001), the interaction between transmission intensity (low vs. high) and asymptomatic infections (HR = 2.43, p-value < 0.0001), the interaction between asymptomatic infections and transformed age (HR = 3.32, p-value <0.0001) and the interaction between asymptomatic infections and year of survey (HR = 1.12, p-value <0.0001). On the other hand, the following covariates were associated with a reduced risk including transmission intensity (low vs. high) though not significant (HR = 0.98, p-value = 0.957), the interaction between transmission intensity (low vs. high) and transformed age though not significant (HR = 0.60, p-value = 0.154), the interaction between transmission intensity (mod-high vs. high) and transformed age (HR = 0.24, p-value < 0.0001), asymptomatic infections (HR = 0.21, p-value < 0.0001), the interaction between transmission (mod-high vs. high) and asymptomatic infections (HR = 0.25, p-value < 0.0001) and year of survey (HR = 0.95, p-value = <0.0001). Finally, among the time varying covariates, the effects of asymptomatic infections

and year of survey were associated with a reduced risk (HR = 0.69, p-value = 0.001 and HR = 0.94, p-value < 0.0001, respectively).

An exploratory analysis showed that every 10-fold increase in asymptomatic parasite density was associated with a 35% increased risk of developing febrile malaria (Hazard Ratio = 1.35, p-value <0.0001). Finally, a sensitivity analysis was conducted using different parasite density thresholds for defining febrile malaria (i.e. any parasitemia in children <1 year and ≥ 2500 parasites/µl for older children) as per previous work (Mwangi *et al.* 2005) to evaluate risk. The effect of asymptomatic infections remained the same as when one cut-off was used to define febrile malaria (i.e. ≥ 2500 parasites/µl for all children) (**Table 2.13**).

Table 2.13. Multivariable analysis to test the effect of different covariates on the risk of developing febrile malaria.

Covariate	Hazard Ratio	Robust Std. Error	z	P>|z|	Confidence Interval	
					Lower	Upper
Main						
Transmission (Low vs. High)	1.01	0.36	0.02	0.984	0.50	2.02
Transmission (Mod-High vs. High)	16.99	5.59	8.61	**<0.0001**	8.91	32.39
Transformed Age	5.31	1.49	5.95	**<0.0001**	3.06	9.21
Transmission (Low vs. High) x Transformed Age	0.59	0.21	-1.51	0.131	0.29	1.17
Transmission (Mod-High vs. High) x Transformed Age	0.23	0.07	-4.70	**<0.0001**	0.13	0.43
Infection Status (Asymptomatic vs. Uninfected)	0.20	0.06	-5.09	**<0.0001**	0.11	0.38
Transmission (Low vs. High) x Infection Status (Asymptomatic vs. Uninfected)	2.44	0.45	4.85	**<0.0001**	1.70	3.49
Transmission (Mod-High vs. High) x Infection Status (Asymptomatic vs. Uninfected)	0.25	0.07	-5.31	**<0.0001**	0.15	0.42
Infection Status (Asymptomatic vs. Uninfected) x Transformed Age	3.42	0.96	4.38	**<0.0001**	1.97	5.94
Year	0.95	0.01	-5.80	**<0.0001**	0.93	0.97
Infection Status (Asymptomatic vs. Uninfected) x Year	1.12	0.02	5.37	**<0.0001**	1.08	1.17
Time-Varying Covariates						
Infection Status (Asymptomatic vs. Uninfected)	0.69	0.08	-3.34	**0.001**	0.56	0.86
Year	0.94	0.01	-7.29	**<0.0001**	0.92	0.95

This model shows the effect of the different covariates on developing febrile episodes with alternative cut-off times of <90 and >90 days and included age-dependent parasite density cut-offs for defining febrile malaria. The p-values in bold represent those that were statistically significant ($p < 0.05$). The symbol x indicates an interaction between the respective covariates. Abbreviations: mod-high - moderate-high.

2.4 Discussion

Asymptomatic infections before the malaria transmission season predicted the risk of developing febrile malaria. This effect was significantly modified by transmission intensity and age. In the moderate to high as well as in the high transmission settings, asymptomatic infections were associated with a reduced risk of febrile malaria in older children (>3-15 years). However, there was no association between asymptomatic infections and the risk in

younger children in the moderate to high and high transmission settings. In contrast, in children of all ages in the lower transmission setting, asymptomatic infections were associated with an increased risk of febrile malaria. Moreover, the effect of asymptomatic infections was modified by time since ascertainment. At low transmission, asymptomatics were at a higher risk than uninfected children until 90 days after ascertainment, following which asymptomatics were at similar risk to uninfected children. On the other hand, at higher transmission and older age, asymptomatics were at a similar risk as uninfected children, until 90 days after ascertainment following which asymptomatics were at a lower risk than uninfected children.

This study is in agreement with studies conducted in high transmission areas where asymptomatic infections have previously been associated with a reduced risk of febrile episodes (Henning *et al.* 2004; Males, Gaye and Garcia 2008; Portugal *et al.* 2017; Buchwald *et al.* 2018), in particular among older children. Additionally, in low transmission areas, some studies have shown that asymptomatic infections were associated with an increased risk of febrile episodes (Henning *et al.* 2004; Njama-Meya, Kamya and Dorsey 2004; Le Port *et al.* 2008; Males, Gaye and Garcia 2008), again consistent with the findings here. Thus, transmission intensity and age modified the effect of asymptomatic infections on the risk of developing febrile malaria.

Children living in malaria-endemic regions with high transmission acquire immunity at a faster rate than those living in low endemic regions as they are repeatedly exposed to infections (Doolan, Dobaño and Baird 2009). Consequently, while immunity to malaria may not necessarily prevent infection, it may help an individual control parasite density and prevent symptoms thus leading to asymptomatic infections (Doolan, Dobaño and Baird 2009). It has been argued that the presence of asymptomatic parasitemia leads to resistance to further infection, a state that is termed “premunition” (Breman 2001). Continuous exposure to malaria infection may continuously prime the immune system and lower the risk of developing febrile malaria. However, the presence of asymptomatic infections implies a risk of parasites going on to cause febrile malaria and may also imply a higher degree of exposure to further infectious bites, and therefore may be associated with a higher risk of febrile

malaria. In this study, older children harbouring asymptomatic infections and living in a region of high transmission intensity, had a reduced risk of developing febrile malaria. On the other hand, younger children with a naïve immune system were at a greater risk of developing febrile malaria if they had asymptomatic infections. The reduced risk with asymptomatic infections becomes more pronounced in older children, which is consistent with a prominent role for acquired immunity, although an additional role of premunition cannot be excluded. Previous studies have indicated that antibodies are predictive of reduced susceptibility to malaria among children with asymptomatic infections but not among uninfected children (Chan *et al.* 2012; Rono *et al.* 2013), implying a role for antibodies in mediating the protection seen in older children with asymptomatic infections. The 90-day period of increased risk following the ascertainment of asymptomatic infections could be due to the increased exposure during to the rainy season. Finally, in the cohorts under study, every 10-fold increase in parasite density was associated with an increased risk of developing a febrile episode. It is possible that higher density asymptomatic infections indicate less host immunity and therefore a greater risk that parasites will evade immune control.

Limitations of this study include the reliance on microscopy diagnosis that may have missed a considerable number of children with sub-microscopic infections since microscopy is less sensitive than PCR (Wu *et al.* 2015). Misclassification of low-density asymptomatic infections as parasite negative would tend to reduce the strength of associations seen, hence this study may be an underestimate of the significance of asymptomatic infections. Malaria transmission is locally heterogeneous (Bejon *et al.* 2010) and hence describing a cohort as uniformly high or low transmission is a simplification, but necessary to study effect modification across groups at different transmission intensity. Episodes of malaria could have been missed during surveillance, either because the episode was self-limiting, or treatment was obtained elsewhere. In mitigation, active surveillance was used and is more sensitive than passive surveillance (Olotu *et al.* 2012). Also, assuming that surveillance is similarly incomplete for all children, a bias arising from this limitation would be unlikely. Host population dynamics such as short-term migration (Prothero 2001) and environmental factors leading to the heterogeneity of risk within cohorts (Bejon *et al.* 2010) were not

accounted for. These may impact on how transmission intensity, age and asymptomatic infections influence the risk of developing febrile malaria.

This study cannot be used to infer the effects of treating asymptomatic infections as it was conducted to study the natural history of asymptomatic infections rather than to make policy recommendations. Furthermore, the findings here may have implications on future studies focussing on asymptomatic infections. More work is needed to establish the role of acquired immunity and premunition in protecting asymptomatic individuals from subsequent febrile episodes. Moreover, there is a need to standardize the definition of asymptomatic infections to ensure that different studies obtain comparable results. For example, future studies should consider including longitudinal follow-up as it will be important when defining asymptomatic infections to exclude pre-symptomatic individuals. However, the identification of biomarkers that can diagnose asymptomatic infections can assist in instances where longitudinal follow ups are lacking.

Chapter 3 : Using deep-sequencing to evaluate *P. falciparum* diversity in asymptomatic and febrile malaria infections

3.1 Introduction

In chapter 3, it was indicated that asymptomatic infections were associated with a reduced risk of febrile malaria episodes (*see section 1.9.3*). Still, the transition from asymptomatic to febrile malaria is yet to be fully understood. While some individuals have been reported to transition due to persistent parasites (Missinou *et al.* 2001; Nsobya *et al.* 2004), a majority of individuals transition due to the introduction of novel parasites (Contamin *et al.* 1996; Babiker *et al.* 1998; Missinou *et al.* 2001; Ofosu-Okyere *et al.* 2001; Kun *et al.* 2002; Nsobya *et al.* 2004; Buchwald *et al.* 2018). These previous findings are based on *msp1*/*msp2*/*glurp* genotyping, the traditional genotyping methods of characterising *P. falciparum* parasite diversity. Such methods however, may fail to generate analysable data due to complexity of infections and may fail to capture minority variant subpopulations.

Amplicon deep-sequencing offers a more sensitive tool compared to traditional genotyping approaches and with its ability to detect minority clones, this chapter sought to examine *(i)* how parasite diversity differs between asymptomatic and febrile malaria infections, *(ii)* how this diversity changes over time and *(iii)* whether episodes of febrile malaria are associated with persistence or the introduction of novel clones compared to those involved in the prior asymptomatic infection.

3.2 Methods

3.2.1 Study design

P. falciparum positive (microscopy and PCR) samples were obtained from the Junju cohort in Kilifi and details regarding this cohort have been mentioned earlier in chapter 3 under section 2.2.1. In this cohort, samples were collected prospectively since 2007 and both asymptomatic and febrile infections have been defined earlier in chapter 2 under section 2.2.2.

3.2.2 DNA extraction

DNA was extracted from frozen blood using the QIAamp DNA Blood Mini Kit (250) (Qiagen, Catalogue 51106) according to the manufacturer's instructions. All samples were equilibrated to room temperature (20–25°C) and all steps were carried out at room temperature. Briefly, DNA extraction involved pipetting 20µl of QIAGEN Protease (or Proteinase K) into the bottom

of a 1.5ml microcentrifuge tube. 200µl of the whole blood sample was added to a microcentrifuge tube and for samples that were less than 200µl, phosphate buffer solution (PBS) was used to top the volume up to 200µl to have similar starting volumes across all samples. 200µl of Buffer AL was then added to each sample and mixed by pulse-vortexing for 15 sec. The samples were then incubated at 56°C for 10 min. 200µl of ethanol (96–100%) was added to each sample, after which the samples were mixed by pulse-vortexing for 15 sec. This mixture was then transferred to a QIAamp Mini spin column (attached to a 2ml collection tube) without wetting the rim. The cap was closed and the spin column centrifuged at 6000 x g for 1 min. The QIAamp Mini spin column was then placed in a clean 2ml collection tube and the tube containing the filtrate was discarded. The QIAamp Mini spin column was uncapped and 500µl Buffer AW1 was added without wetting the rim. The cap was then closed and the spin column centrifuged at 6000 x g for 1 min. The QIAamp Mini spin column was placed in a clean 2ml collection tube and the tube containing the filtrate was discarded. The QIAamp Mini spin column was uncapped and 500µl Buffer AW2 was added without wetting the rim. The cap was then closed and centrifuged at full speed (20,000 x g) for 3 min. The QIAamp Mini spin column was placed in a new 2ml collection tube and the old collection tube with the filtrate was discarded. The spin-column was then centrifuged at full speed for 1 min to eliminate the chance of possible Buffer AW2 carry-over. The QIAamp Mini spin column was placed in a clean 1.5ml microcentrifuge tube and the collection tube containing the filtrate was discarded. The QIAamp Mini spin column was uncapped and 50µl Buffer AE was added. The spin-column was incubated at room temperature (20–25°C) for 1 min to increase the DNA yield and then centrifuged at 6000 x g for 1 min. At this step, the eluate was ready for PCR amplification.

3.2.3 PCR amplification

P. falciparum apical membrane antigen 1 (*ama1*, PF3D7_1133400) amplicons spanning the nucleotides 465-923 were generated from each sample, in duplicate, using primers designed in this study (**Table 3.1**). This region of the *ama1* gene has been found to have the highest nucleotide diversity (Polley and Conway 2001) and hence would serve as a good marker to explore parasite diversity.

Table 3.1. List of primers for PCR amplification

Primer	Sequence '5-3'	Length (bases)
AMA1_F	GAA ATG TCC AGT ATT TGG TAA AGG	24
AMA1_R	CCC ATA ATC CGA ATT TTG CAT TC	23

PCR included the following steps: initial denaturation (98°C – 30 sec), followed by 30 cycles of denaturation (98°C – 10 sec), annealing (60°C for 30 sec for all genes), extension (72°C – 30 sec), and final extension (72°C – 2 min).

For PCR, the following were used: 1μl of template DNA (final concentration < 50ng), 0.2μl of Q5® High-Fidelity DNA Polymerase (final concentration 0.02U/μl, New England BioLabs), forward primers tagged with Roche® multiplex identifiers (MIDs, **Table 3.2**) that enabled the demultiplexing of sequence reads into separate sequence reads for each sample (1μl, final concentration 10mM), reverse primers (1μl, final concentration 10mM), 0.4μl of 10mM dNTPs (final concentration 10μM, Bioline), 4μl of 5X Q5 reaction buffer (final concentration 1X), 12.4μl of nuclease-free water and the PCR conditions indicated in (**Table 3.1**).

Table 3.2. 10-base multiplex identifier (MID) set sequences

MID Tag	Sequence	MID Tag	Sequence
MID01	ACGAGTGCGT	MID14	TCACGTACTA
MID02	ACGCTCGACA	MID15	CGTCTAGTAC
MID03	AGACGCACTC	MID16	TCTACGTAGC
MID04	AGCACTGTAG	MID17	TGTACTACTC
MID05	ATCAGACACG	MID18	ACGACTACAG
MID06	ATATCGCGAG	MID19	CGTAGACTAG
MID07	CGTGTCTCTA	MID20	TACGAGTATG
MID08	CTCGCGTGTC	MID21	TACTCTCGTG
MID09	TCTCTATGCG	MID22	TAGAGACGAG
MID10	TGATACGTCT	MID23	TCGTCGCTCG
MID11	CATAGTAGTG	MID24	ACATACGCGT
MID12	CGAGAGATAC	MID25	ACGCGAGTAT
MID13	ATACGACGTA	MID26	ACTACTATGT

Amplicons were generated using forward primers tagged on the 5' end with the above listed MIDs. This enabled the unique identification of sequences from each sample during sequence data demultiplexing.

The final reaction volume was 20μl and the PCR amplification products were visualised on 1% agarose gels stained with RedSafe™ Nucleic Acid Staining Solution (iNtRON Biotechnology DR). PCR negative samples were taken through a second and final round of PCR with 1.5μl of

template DNA. Also included for sequencing were lab controls comprising of a known mixture of two *P. falciparum* isolate: 3D7 (MRA-102G, BEI Resources, Manansas, VA) and Dd2 (MRA-150G, BEI Resources). These two parasite isolates differ since they carry the wild-type and mutant *ama1* and *mdr1* variants, respectively, hence were appropriate sequencing controls. From this control mixture, *ama1* amplicons were generated from six replicates each of the following percentage ratios of 3D7 to Dd2: 50:50, 75:25, 85:15, 95:5 and 100:0.

3.2.4 PCR product purification

This protocol was applied to the amplified PCR products to remove of enzymes, nucleotides, primers and PCR buffer components as these may interfere with sequencing. The Zymo ZR-96 DNA Clean & Concentrator-5 Kit (Zymo Research, Catalog Number D4024) was used and all steps were performed at room temperature (20-25°C) and centrifugation was performed between 3,000 - 5,000 x g at room temperature.

Purification of the PCR products involved adding 2-7 volumes of DNA Binding Buffer to each volume of PCR product and the mixture was vortexed briefly. The mixture was then transferred to the wells of a Zymo-Spin™ I-96 Plate mounted on a collection Plate. Centrifugation was then conducted for 5 min until the sample mixtures were completely filtered, after which the flow-through was discarded. 300μl DNA Wash Buffer was then added to each well of the Zymo-Spin™ I-96 Plate and the mixture was centrifuged for 5 min. This was repeated once more, but centrifugation was done for 15 min to ensure that the DNA Wash Buffer was eliminated from the sample. 30μl DNA Elution Buffer was added to the column matrix in each well and the Zymo-Spin™ I-96 Plate was then transferred onto an Elution Plate and centrifuged for 3 min to elute the DNA. Finally, the DNA was quantified using Quant-iT™ dsDNA Assay Kit, High Sensitivity (Invitrogen). At this stage, the ultra-pure DNA was ready for library preparation.

3.2.5 Sequencing library preparation

This protocol was performed according to the manufacturer's instructions using The KAPA HyperPrep Kit which comprised 3 products: KAPA Library Amplification Primer Mix (10X, Roche Catalogue Number 07962363001 and Kit Code KK8504), KAPA Dual-Indexed Adapter

Kit (15 μM, Roche Cat. No 08278555702, Kit Code KK8722) and KAPA Adapter Dilution Buffer (25 mL, Roche Cat. No 08278539001, Kit Code KK8721). Also included were magnetic Agencourt AMPure XP beads (60 mL, Beckman Coulter, Cat. No A63881). The essence of this protocol was to attach KAPA adapters to the ends of each amplicon to enable them to bind to a sequencing flow cell and enable sequencing of each DNA fragment. Briefly, the quantified PCR products were normalised by diluting in EB Buffer (Qiagen) to achieve equimolar concentrations of 1ng each across all samples. The equimolar amplicons were then mixed to come up with amplicon pools of non-overlapping MIDs and since 26 unique MIDs were used, this allowed up to 13 samples to be sequenced in duplicate in each sequencing library. These pooled amplicons were then taken through library preparation which involved attaching dual index adapters using the KAPA Dual-Indexed Adapter Kit (KAPA BioSystems) and the KAPA Hyper Prep Kit (KAPA BioSystems).

Library preparation commenced by taking the PCR products through an end-repair and A-tailing process, which involved creating blunt ends and adding an A-tail to the 5' and 3' ends of each amplicon. In turn, this created binding sites for the Illumina-compatible KAPA adapters. By incubating the products from above with the KAPA adapters, the adapters were ligated to the DNA fragments. Post ligation cleanup was performed using the AMPure beads to size-select for adapter-ligated *ama1* amplicons (636bp, including 516bp of *ama1* and 120 of KAPA adapters) and this eliminated free adapters, primers, nucleotides, salts and DNA fragments not ligated to adapters. Library amplification was then performed to increase the concentration of the adapter-ligated DNA fragments in a thermal cycler and a final round of clean up was performed with the magnetic AMPure beads as described above. Adapter ligation was confirmed by running a random sample of 20 libraries on the Agilent 2200 TapeStation System using the Agilent High Sensitivity D1000 ScreenTape System (5067-5584). Since adapter ligation increased the length of amplicons by 120bp, a peak of approximately 636bp was expected on the TapeStation. Finally, the *ama1* adapter ligated amplicon libraries were quantified, normalised to 1ng each and mixed to form one final pool.

3.2.6 Amplicon deep sequencing

Paired end sequencing (2x300 bp chemistry) of the final pool was performed on the MiSeq platform (Illumina) using the MiSeq Reagent Kit v3 (Illumina) in two labs: a subset of 127 asymptomatic samples were amplified and sequenced at the KEMRI-Wellcome Trust Research Programme (KWTRP) lab while 408 asymptomatic samples and 114 febrile malaria samples were amplified and sequenced at the Infectious Disease Epidemiology and Ecology Lab (IDEEL) at the University of North Carolina (Chapel Hill) (**Figure 3.1**).

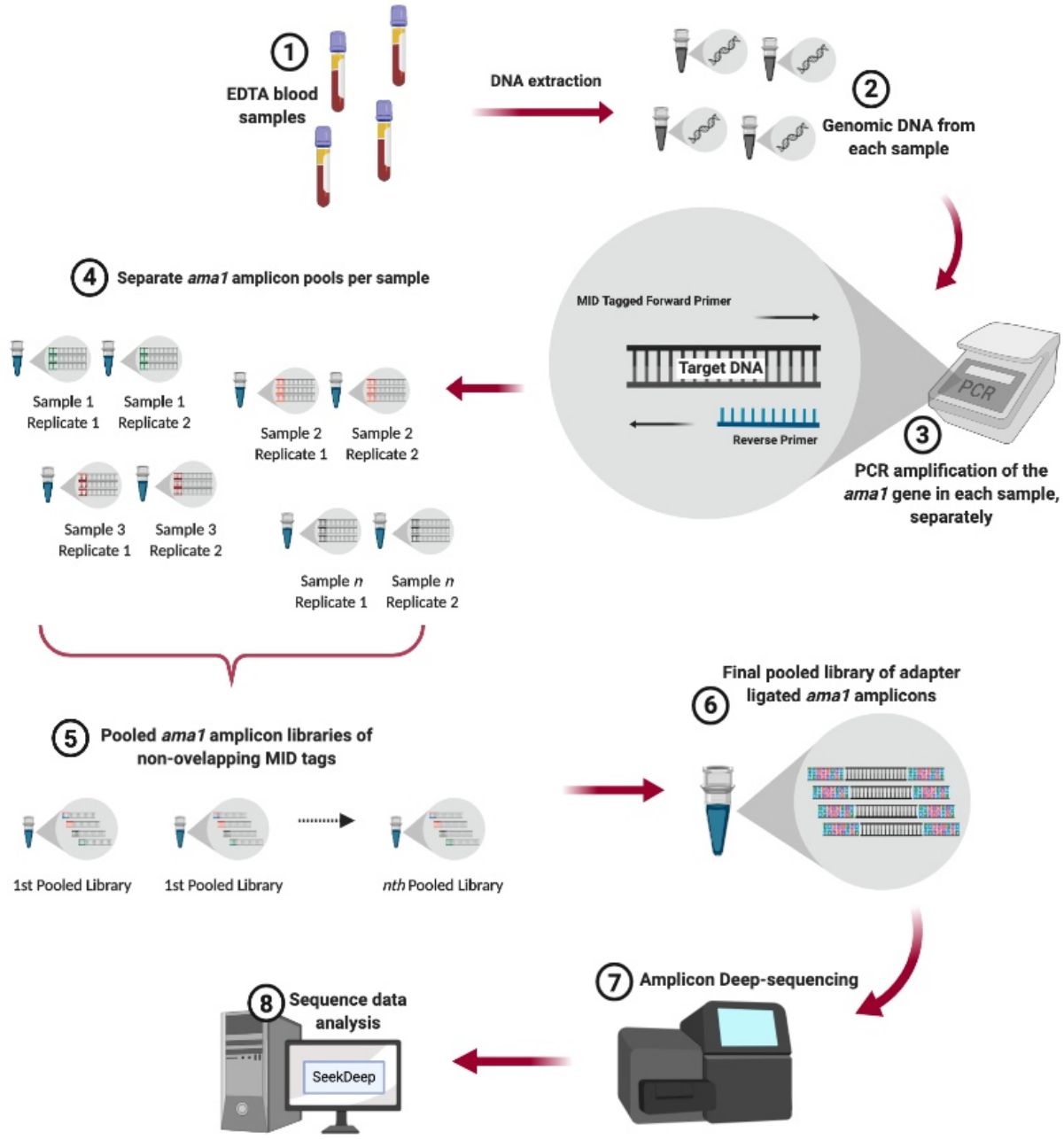


Figure 3.1. Schematic representation of the deep-sequencing pipeline. *(1)* the pipeline began with the extraction of DNA from frozen EDTA blood so as to generate *(2)* separate DNA pools from each sample. From the extracted DNA, *(3)* PCR amplification of the *ama1* gene was conducted using MID-tagged forward primers and untagged reverse primers and this generated *(4)* separate pools of amplicons from each sample, in duplicate. *(5)* The amplicons were then grouped into pools of non-overlapping MIDs from which sequence libraries were generated by attaching library specific adapters and sequencing adapters to the 5' and 3' ends of each amplicon. The library specific adapters and unique MIDs enabled the demultiplexing of sequence data to facilitate the identification of samples from which each sequence originated. *(6)* A final pool was created containing equimolar amounts of each sequencing library and this final pool was *(7)* sequenced on the MiSeq Illumina platform. *(8)* The sequence data generated was analysed using the SeekDeep software.

3.2.7 Sequence and data analysis

Sequence analysis was performed on SeekDeep v2.6.4 (Hathaway *et al.* 2018). In brief, three of SeekDeep programs were used to: ***(i)*** extractorPairedEnd – an algorithm that carries out sequence demultiplexing and read filtering, enabling sequence data to be separated into individual samples based on the MIDs. Also, the tagged forward primer reads and untagged reverse primer reads were combined to generate the original amplicon fragments by sequence matching at the regions where the two reads overlap. ***(ii)*** qluster - an algorithm that carries out DNA sequence clustering and frequency estimation of the variants per demultiplexed data subset. *ama1* haplotypes were generated based on the amino acid residues from all the polymorphic codons of the translated DNA sequences ***(iii)*** processClusters - an algorithm that carries out multiple PCR replicate comparisons to determine if the DNA sequence variants were identified in all sample replicates, as well as carrying out a comparison of the DNA variant frequencies across all samples. DNA variants that did not occur in two sample replicates were discarded. Also, since one of the ratios of the lab sequencing controls included a minor variant at 5% frequency (3D7 95% vs. Dd2 5%), DNA variants that were detected below this frequency were discarded. However, if such variants were detected above 5% frequency in other samples, they were retained. So as to reduce the rate of calling false positives, DNA variants that were detected in only one sample replicate were also discarded. Chimeric reads that originate as PCR artefacts were also excluded. Such reads are generated when the extension of an amplicon is terminated, and the aborted product functions as a primer in the next PCR cycle, thereby synthesizing a single sequence sourced from two different templates (Haas *et al.* 2011).

3.2.8 Statistical analysis

Statistical analysis was carried in R v3.6.0 (R Core Team ,2014). COI was defined as the number of DNA variants in each sample and the Wilcoxon signed-rank test was used to compare the difference in COI between asymptomatic and febrile malaria samples using the R package, ggpubr v0.2 (Kassambara 2020). The Poisson regression analysis was used to examine the impact of age and parasitemia on COI, while a logistic regression was used to test the impact of COI on time to first febrile malaria episode based on data from a previous study (Wamae

et al. 2019). For the first-febrile episode with "*x*" number of persistent *ama1* variants and sharing a single variant of prevalence *"y"*, the binomial probability that this variant is found by chance in a subsequent infection was calculated as $1-(1-y)^x$. This probability reflected the likelihood that the persistent genotype represented a new mosquito-bite infection. In cases where there were multiple shared persistent variants, the combined probability that all shared variants would appear in a reinfection was calculated as the product of the individual probabilities (Lin *et al.* 2015). All estimates were reported as incidence rate ratios (IRR) with confidence intervals (C.I) and all plots were generated using the R package, ggplot2 v3.1.1 (Wickham 2016).

3.3 Results

3.3.1 Reproducibility and sensitivity analysis based on control mixtures

A total of 49.45 million reads (combined forward and reverse) were generated and 26.2 million (52.9%) of these were discarded as the MID tags could not be identified. A further 6.2 million (12.5%) were discarded as the primer sequences could not be detected, 9 million (18.2%) were discarded because the pair reads did not overlap, 55,000 (0.1%) were discarded because they were identified as contaminated, possibly from bacterial DNA. Finally, 1 million (2%) reads were discarded because they were either chimeric, represented variants that did not occur in two sample replicates. Therefore, a total of 6.9 million (13.9%) reads were retained and used for analysis.

A total of 30 individual lab controls were sequenced, representing six replicates of 3D7 and Dd2 mixtures, with one failure (**Table 3.3**). The median read depth for detecting DNA variants in the lab controls was 2,877 reads (range 2,109 – 6,477). Both 3D7 and Dd2 *ama1* DNA sequences were identified across all control mixtures with frequencies similar to those expected and **Table 3.3** shows the frequencies of these sequences in each control replicate.

Table 3.3. Frequencies of known *ama1* DNA sequences in sequencing controls

Control Set	Replicate	Lab Isolate	DNA sequence frequency [%]	
			Expected	Observed
Set 1	Replicate 1	3D7	50	59
		Dd2	50	41
	Replicate 2	3D7	50	58
		Dd2	50	42
	Replicate 3	3D7	50	56

		Dd2	50	44
	Replicate 4	3D7	50	57
		Dd2	50	43
	Replicate 5	3D7	50	56
		Dd2	50	44
	Replicate 6	3D7	50	*Failed Sequencing*
		Dd2	50	
Set 2	Replicate 1	3D7	75	71
		Dd2	25	29
	Replicate 2	3D7	75	72
		Dd2	25	28
	Replicate 3	3D7	75	73
		Dd2	25	27
	Replicate 4	3D7	75	71
		Dd2	25	29
	Replicate 5	3D7	75	71
		Dd2	25	29
	Replicate 6	3D7	75	73
		Dd2	25	27
Set 3	Replicate 1	3D7	85	84
		Dd2	15	16
	Replicate 2	3D7	85	82
		Dd2	15	18
	Replicate 3	3D7	85	83
		Dd2	15	17
	Replicate 4	3D7	85	84
		Dd2	15	16
	Replicate 5	3D7	85	85
		Dd2	15	15
	Replicate 6	3D7	85	82
		Dd2	15	18
Set 4	Replicate 1	3D7	95	94
		Dd2	5	6
	Replicate 2	3D7	95	96
		Dd2	5	4
	Replicate 3	3D7	95	94
		Dd2	5	6
	Replicate 4	3D7	95	94
		Dd2	5	6
	Replicate 5	3D7	95	95
		Dd2	5	5
	Replicate 6	3D7	95	94
		Dd2	5	6
Set 5	Replicate 1	3D7	100	100
		Dd2	0	0
	Replicate 2	3D7	100	100
		Dd2	0	0
	Replicate 3	3D7	100	100
		Dd2	0	0
	Replicate 4	3D7	100	100
		Dd2	0	0
	Replicate 5	3D7	100	100
		Dd2	0	0
	Replicate 6	3D7	100	100
		Dd2	0	0

Mixtures of two lab controls (3D7 and Dd2) were prepared by mixing DNA from the two isolates in the ratios indicated under the "*Expected*" column, six replicates each. The "*Observed*" column represents

the frequencies at which each isolate was detected upon sequencing. Sequencing failed for replicate 6, under Control Set 1.

Table 3.4 shows the comparison between the mean expected DNA sequence frequencies versus the observed frequencies. Overall correlation between expected vs. observed frequencies was high (Pearson correlation = 0.89 and p-value < 0.001), hence, analysis of field samples could proceed.

Table 3.4. Mean frequencies of known DNA sequences in sequencing controls

Control Set	Lab Isolate	Mean DNA sequence frequency [%]	
		Expected	Observed
Set 1	Dd2	50	57
	3D7	50	43
Set 2	Dd2	25	28
	3D7	75	72
Set 3	Dd2	15	17
	3D7	85	83
Set 4	Dd2	5	5
	3D7	95	95
Set 5	Dd2	0	0
	3D7	100	100

Mixtures of two lab controls (3D7 and Dd2) were prepared by mixing DNA from the two isolates in the ratios indicated under the "*Expected*" column, six replicates per control set. The "*Observed*" column represents the mean frequencies at which each isolate was detected upon sequencing.

3.3.2 PCR amplification and deep-sequencing of samples

A total of 535 asymptomatic and 114 febrile samples were included in this study. This set included 114 asymptomatic samples paired with the participant's first febrile malaria sample (**Table 3.5**).

Table 3.5. The number of samples across asymptomatic and febrile infections per year

Asymptomatics (Cohort)	Febrile (Cohort)
Sampling Year (n)	Sampling Year (n)
2007 – 55	-
2008 – 110	-
2009 – 96	2009 – 20
2010 – 223	2010 – 39
2011 – 27	2011 – 28
2012 – 2	2012 – 4
2013 – 3	2013 – 4
2014 – 16	2014 – 15

2015 – 4 2016 – 1	2015 – 3 2016 – 1
Age Range 3 months – 14 years	Age Range 8 months – 13 years
Parasites/μL (Microscopy) Median 128 (range 0 – 600,482)	Parasites/μL (Microscopy) Median 26,260 (range 2,560 – 550,000)
Parasites/μL (PCR) Median 503 (range 0.2 – 2,069,633)	Parasites/μL (PCR) -
Gender Females – 266 Males – 271	Gender Females – 54 Males – 60

Entries marked with "-" under sampling year indicate time-points that were not included in the study, while for parasites/μL (PCR), this indicates data was not available. Microscopy negative asymptomatic samples were positive by PCR.

177/535 asymptomatic, 98/114 febrile, 30/114 paired asymptomatic-febrile malaria samples had good quality data for analysis. The median read depth for detecting DNA variants in asymptomatic and febrile malaria samples was 8,941 reads (range 427 – 199,885) and 3,934 reads (range 810 – 8,032), respectively. For the samples that were successfully sequenced, the median age for asymptomatic participants was 6.9 years (range 1-12 years), 7.5 years (range 9 months – 13 years) for participants with febrile malaria and 7 years (range 1-11 years) for the participants with the paired asymptomatic-febrile malaria samples (**Table 3.6**).

Table 3.6. Summary of PCR and sequencing success for asymptomatic and febrile malaria samples.

Year	Total	Failed PCR	Failed Sequencing	Successfully Sequenced
Asymptomatics				
2007	55	29	2	24
2008	110	49	17	44
2009	96	51	18	27
2010	223	140	14	69
2011	27	16	4	7
2012	2	0	0	2
2013	3	2	0	1
2014	16	10	1	5
2015	4	3	1	0
2016	1	1	0	0
Symptomatics				
2009	20	1	1	18
2010	39	1	2	36
2011	28	3	1	24
2012	4	0	1	3
2013	4	0	1	3
2014	15	0	0	15
2015	3	0	1	2

2016	1	0	0	1

Failed PCR represents samples that were unsuccessfully amplified while failed sequencing represents samples that were unsuccessfully sequenced.

Table 3.7 shows the summary of the distribution of parasitemia by microscopy and PCR for the samples that were successfully sequenced and those that failed. Overall, the geometric mean parasitemia per microliter was higher in samples that were successfully sequenced compared to those that failed. Moreover, 207/535 (38%) asymptomatic samples that failed sequencing were positive by PCR but negative by microscopy.

Table 3.7. Distribution of parasitemia by microscopy and PCR in the samples included in this study

Samples	Parasite Positive	PCR/Sequencing	Microscopy and PCR Parasitemia/μl		
			Number of Samples/n	Geometric Mean	Range
Asymptomatics	Microscopy + PCR	Successful	22/177	1,500	40 - 600,482
		Failed	153/358	802	40 - 2,200,000
	PCR Only	Successful	155/177	916	22 - 23,625
		Failed	207/358	35	0.2 - 11,120
Febrile	Microscopy Only	Successful	96/122	50,753	2,560 - 1,280,000
		Failed	26/122	24,457	2,600 - 510,000

*PCR data was not available for febrile malaria samples. Under "number of samples/n", the denominator 177 and 358 represent samples that were successfully and unsuccessfully sequences, respectively.

There was a high correlation between number of samples sequenced and variants of DNA sequences detected (Pearson correlation = 0.96, p-value <0.0001, **Table 3.8**).

Table 3.8. Total *ama1* variants detected in asymptomatic and febrile infections over time

Sampling Year	Asymptomatic Infections		Febrile Infections	
	Total Samples	*ama1* variants	Total Samples	*ama1* variants
2007	24	29	-	-
2008	44	49	-	-
2009	27	34	18	21
2010	70	59	36	40
2011	6	7	24	27
2012	2	4	3	5
2013	1	1	3	5
2014	5	7	15	22
2015	-	-	2	4
2016	-	-	1	1

This table compares the number of samples successfully sequenced per year with the number of *ama1* variants detected. In hyphens "-" are time-points that were not successfully sequenced for asymptomatic infections and those that were not sampled for febrile infections.

A total of 43 SNPs were detected in the amplified region of the *ama1* gene. In turn, 34 polymorphic codons were used to generate the list of haplotypes including codons: 162, 167, 172, 173, 175, 176, 187, 189, 190, 196, 197, 199, 200, 203, 204, 206, 207, 224, 225, 227, 228, 230, 243, 244, 245, 267, 269, 282, 283, 285, 286, 296, 299 and 300. 23 *ama1* variants were found to occur in only one sample each, including 13 in asymptomatic and 10 in febrile malaria infections, and these were excluded from subsequent analysis. Additionally, 63 samples (43 asymptomatic and 21 febrile) were found to harbour variants with frequencies <5% (38 variants in total). However, these variants were found to occur at frequencies ≥5% in other samples and were included in subsequent analysis. Consequently, a total of 72 variants (68 haplotypes) were identified. 68 of these variants (64 haplotypes) were detected in asymptomatic samples, 51 variants (47 haplotypes) were detected in febrile malaria samples and 48 of these (44 haplotypes) occurred in both asymptomatic and febrile malaria samples (**Table 3.9**).

Table 3.9 The frequencies of *ama1* variants detected *(i)* across all samples, *(ii)* within asymptomatic and *(iii)* within febrile infections

Variant ID	Frequency [n]			Haplotype
	Combined	Asymptomatic	Febrile	
PFAMA1.00	9.9 [77]	13.8 [61]	6.3 [16]	KTENDNPMNGRDLKNEYMNNEDKDKEKKSQNDEE
PFAMA1.01	6.9 [56]	5.3 [33]	11.9 [23]	NTENDKLMDQRHFKDKYMINKDKDKEKISQNDEK
PFAMA1.68	0.2 [2]	0.3 [2]	-	
PFAMA1.02	5.6 [43]	5.8 [29]	7.2 [14]	NKGNDELIDDRDFKNEYMNNQYEYKQKISQNDEK
PFAMA1.03	2.9 [39]	1.8 [16]	-	NTGNYKLMDDRLLKDEDMNNKYNDKEKKSQNDEE
PFAMA1.04	7.8 [35]	0.5 [3]	-	NTGNYELMDERHFKDKYMINKDKDKEKISQNDEK
PFAMA1.54	0.6 [3]	1 [3]	-	
PFAMA1.05	4.6 [35]	4 [24]	7.1 [11]	NTENDNLINGKDFKDEDMNKEYEDNEKKSQNDEK
PFAMA1.06	2.5 [29]	2.6 [18]	3.1 [11]	NTGKDNLINGRDLKNEDMNNKDKNKQKKSQNDEK
PFAMA1.37	0.7 [6]	0.7 [4]	1.1 [2]	
PFAMA1.07	3 [28]	3.1 [20]	3.7 [8]	KTENDELMDDRRFKDEYMNNKYNDKEKKSQNDEK
PFAMA1.08	2.7 [28]	4.6 [26]	0.2 [2]	NTENDELMDRRDFKNEDMNKEYKDKEKKSQNDEK
PFAMA1.09	3.2 [26]	4.4 [20]	2.1 [6]	NTENDKLMDQRHFKDKYMINKDKDKQIISQNDEK
PFAMA1.10	2.8 [24]	1.9 [15]	5.3 [9]	NTENDKHMDDRLLKDEDMINKDKDKQKKSQDDEK
PFAMA1.11	2 [20]	2 [13]	2.7 [7]	NTENDKLMDQRHFKDEDMINKDKDKEKKLQNHEE
PFAMA1.12	2 [18]	3 [16]	0.9 [2]	KTENDNPMNGRDLKNEYMNNKYNDNEKKSQNDEK
PFAMA1.13	2.2 [15]	2.9 [13]	1.6 [2]	NTGKDNLINGRDLKNEDMNNKYNDKQKKLENDEE
PFAMA1.14	1.4 [15]	2.1 [13]	0.8 [2]	NTENDELMDQRHFKDKYMINKDKDKQIISQNDEK
PFAMA1.15	1.4 [15]	1.2 [9]	2.3 [6]	KTENDNPMNGRDLKNEYMNNKYNDNEKKSQNDEE
PFAMA1.16	0.8 [14]	1.3 [11]	0.4 [3]	KTENDNPMNGRDLKNEYMNNEDKDKEKKSQNDEK
PFAMA1.17	1.4 [13]	0.6 [7]	3.4 [6]	NKGNDNLINGRDFKNEYMNNKYNDKQKKLENDEE
PFAMA1.18	1.3 [13]	1.9 [11]	0.7 [2]	NTGKDNLINGKDFKDEDMNKEYKDKEKKSQNDEK

PFAMA1.19	1.2 [11]	1.4 [8]	1.3 [3]	
PFAMA1.20	1.6 [10]	1.7 [8]	1.9 [2]	KTENDELIDQRHLKDEYMNNKYEDKEKKLENDEE
PFAMA1.21	1 [10]	1.7 [10]	-	NKGNDELIDDRDFKNEYMNNQYEDKQKKSQNHEK
PFAMA1.22	0.9 [10]	0.8 [8]	1.4 [2]	NTENDELMDRRDFKNEDMNNKDKNKQKKSQNDEE
PFAMA1.23	0.7 [10]	1.1 [8]	0.1 [2]	NTGNYEHMDERHFKDKYMINKYEDKQKKLQNDEE
PFAMA1.24	1 [9]	1.7 [7]	0.1 [2]	NTENYELMDERHFKDKYMINKDKDKEKISQNDEK
PFAMA1.25	1 [9]	0.6 [5]	1.9 [4]	NKGNDELINGRDFKNEYMNKEDKDKEKKSQNDEK
PFAMA1.26	0.8 [9]	1 [8]	0.8 [1]	NTENDNLINGRDLKNEDMNNKDKNKQKKSQNDEK
PFAMA1.27	1.2 [8]	1.4 [5]	1.3 [3]	KTENDKLMDQRHFKDKYIINKDKDKEKISQNDEE
PFAMA1.28	0.5 [8]	0.8 [6]	0.2 [2]	NKGNDNLINGRDFKNEYMNNEDKDKEKKSQNDEK
PFAMA1.29	0.4 [8]	0.6 [5]	0.2 [3]	NKGNDNLINGRDFKNEYMNNKYEDKQKKLQNHEE
PFAMA1.30	0.9 [7]	0.8 [5]	1.6 [2]	KTENDELMDDRRFKDEYMNNKYNDKEKKLENDEE
PFAMA1.31	0.8 [7]	0.6 [5]	1.4 [2]	NTENYELMDERHFKDKYMINKDKDKEKKSQNDEE
PFAMA1.32	0.8 [7]	1 [4]	0.8 [3]	NTENDKLMDERHFKDKYMINKDKDKEKISQNDEK
PFAMA1.33	0.7 [7]	0.03 [1]	2.1 [6]	NTENDKLIDQRDLKNEYMNNEYKDKQKKLENDEK
PFAMA1.34	0.6 [7]	0.6 [6]	1 [1]	KTENDELIDDRDSKDEYMNNEYKDKEKKSQNDEK
PFAMA1.35	1 [6]	1.7 [5]	0.03 [1]	NTGKDNLINGRDLKNEDMNNKDKNKQKKLQNDEE
PFAMA1.36	0.8 [6]	0.8 [4]	1.1 [2]	NKGNDKLIDDRDFKNEYMNKEYKDKEKKSQNHEK
PFAMA1.38	0.7 [6]	1.1 [6]	-	NKGNDNLINGRDFKNEYMNNKYNDKQKKLQNHEE
PFAMA1.39	0.4 [5]	0.7 [5]	-	NTENDKLMDQRDFKNEYMNNEYKDKQKKLENHEK
PFAMA1.40	0.4 [5]	0.1 [4]	1 [1]	KTENDKLMDDRDLKNEYMNNKYNDNEKKSQNDEK
PFAMA1.41	0.3 [5]	0.6 [5]	-	NTGNYEHMDERHFKDKYMINKYEDKQIISQNDEK
PFAMA1.42	0.3 [5]	0.3 [4]	0.4 [1]	NTENDNLINGKDFKDEDMNKEYEDNEKKSQNDEE
PFAMA1.43	1.1 [4]	-	-	NKGNDNLIDHRDFKNEYMNNKYNDKEKKSQNDEK
PFAMA1.44	0.9 [4]	1 [3]	1 [1]	KTENDELMDDRRFKDEYMNNKYNDNEKKSQNDEK
PFAMA1.45	0.8 [4]	0.4 [1]	1.7 [3]	NTGNYELMNGRDLKNEDMNNKDKNKQKKSQNDEK
PFAMA1.46	0.7 [4]	0.8 [3]	0.9 [1]	NKGNDKLIDDRDFKNEYMNNQYEDKQKKLQNHEK
PFAMA1.47	0.4 [4]	-	1.2 [4]	NKGNDELIDDRDFKDEDMNKEYKDKEKKSQNDEK
PFAMA1.48	0.3 [4]	0.3 [3]	0.5 [1]	NKGNDNPMNGRDLKNEYMNNQDKDKEKKSQNDEE
PFAMA1.49	0.3 [4]	0.5 [4]	-	NTENDNPMNGRDLKNEYMNNEDKDKEKKSQNDEE
PFAMA1.50	0.3 [4]	0.5 [4]	-	NKGNDELIDDRDFKNEYMNNQYEDKQKKLQNDEE
PFAMA1.51	0.2 [4]	0.4 [4]	-	NTVNDKLMDDRLLKDEDMINKDKDKEKKLENDEK
PFAMA1.52	0.1 [4]	0.1 [2]	0.1 [2]	KTENDNPMNGRDLKNEYMINKDKDKEKKLENDEE
PFAMA1.53	0.7 [3]	0.6 [1]	1.2 [2]	KTENDNPMNGRDLKNEDMNNKDKNKQKKSQNDEK
PFAMA1.55	0.4 [3]	0.7 [3]	-	NTGNYEHMDERHFKDKYMINKYEDKQKKLQNDEK
PFAMA1.56	0.4 [3]	0.6 [3]	-	NKGNDELIDQRHFKDEYMNNKYEDKQKKSQNDEK
PFAMA1.57	0.3 [3]	0.6 [3]	-	NKGNDELIDDRDFKNEYMNNQYEDKQIISQNDEE
PFAMA1.58	0.6 [2]	0.6 [1]	1 [1]	NTENDNLINGKDFKDEDMNKEYKDKEKKSQNDEK
PFAMA1.59	0.6 [2]	-	2 [2]	NTENDNLIDHRDSKNEYMNNQYKDKQKKSQNDEK
PFAMA1.60	0.5 [2]	0.8 [2]	-	KTENDKLMDRRHLEDEYMNNEYKDKEKKLENDEK
PFAMA1.61	0.4 [2]	0.2 [1]	0.9 [1]	NTENDKHMDDRVLKDEYMNNEDKDKQKKSQNDEK
PFAMA1.62	0.4 [2]	0.6 [1]	0.1 [1]	NKGNDNLMNGRDFKNEYMNNKYNDKEKKSQNDEK
PFAMA1.63	0.3 [2]	0.6 [2]	-	NTENDKLIDQRDLKNEYMNNKYNDKQKKLENDEE
PFAMA1.64	0.3 [2]	-	1 [2]	NKGNDNLMNGRDFKNEYMNNKYEDKQKKLQNHEE
PFAMA1.65	0.3 [2]	0.5 [1]	1 [1]	NTENDKLIDQRDLKNEYMNNEYKDKQKKLENHEK
PFAMA1.66	0.3 [2]	0.6 [2]	-	NTGKDELIDDRDFKNEYMNNQYEDKQKKSQNHEK
PFAMA1.67	0.2 [2]	0.4 [2]	-	NKGNDELIDDRDFKNEYMNNKYEDKQKKLQNHEK
PFAMA1.69	0.1 [2]	0.01 [1]	0.4 [1]	NTGNYKLMDDRLLKDEDMINKDKDKEKKLENDEK
PFAMA1.70	0.1 [2]	0.1 [2]	-	NKGNDELIDQRHFKDEYMNNKYEDKQKKLQNHKK

PFAMA1.7	0.04 [2]	0.1 [2]	-	NTGKDNLINGKDFKDEDMNKEYEDNEKKSQNDEK

The variable "combined" represents frequencies across both asymptomatic and febrile infections while n represents the number of samples. The entries marked in grey indicate entries where two unique DNA variants had synonymous SNPs and collapsed into the same amino acid haplotype. Entries marked with "-" represent variants that were not detected in the respective cases.

Of all the samples sequenced, the median read depth in the asymptomatic samples sequenced at the KWTRP and IDEEL lab was 28,347 reads (range 895 – 199,885) and 4,947 reads (range 427 – 29,398), respectively. Of the 177 asymptomatic samples that were successfully sequenced, 88 were successfully sequenced at the KWTRP lab while 89 were successfully sequenced at the IDEEL lab. Of the 68 *ama1* variants detected in the successfully sequenced asymptomatic samples, 45 (66.1%) variants overlapped between the two labs, while 15 and 8 were detected only at KWTRP and IDEEL labs, respectively.

Compared to SSA *ama1* sequences that span the region that was amplified in this study, 61/68 haplotypes were found to occur in a dataset of 328 haplotypes from Kilifi (Osier *et al.* 2010 and unpublished data), 16/68 haplotypes were found in a dataset of 35 variants from Nigeria (Polley and Conway 2001), 22/68 haplotypes were found in a dataset of 122 variants from Mali (Takala *et al.* 2009) and 19/68 haplotypes were found in a dataset of 50 variants from The Gambia (Tetteh *et al.* 2009). The other haplotypes not identified in other regions appear to be unique to Kilifi.

From the combined asymptomatic and febrile malaria infections data, the dominant *ama1* variant (PFAMA1.00) was detected in 77/273 samples with a population frequency of 9.9% in the asymptomatic and febrile infections, combined. The least dominant *ama1* variant (PFAMA1.71) was detected in two individuals and had a population frequency of 0.04% in the asymptomatic and febrile infections, combined (**Table 3.9**). Within asymptomatic samples only, the most dominant *ama1* variant (PFAMA1.00) was detected in 61/177 individuals with a population frequency of 13.8% while the least dominant variant (PFAMA1.69) was detected in two individuals with a population frequency of 0.01% (**Table 3.9**). Within the febrile malaria samples only, the most dominant variant (PFAMA1.01) was detected in 23/98 individuals with

a population frequency of 11.9%, while the least dominant variant (PFAMA1.35) was detected in two individuals with a population frequency of 0.03% (**Table 3.9**).

The median COI observed in asymptomatic and febrile malaria infections was 3 (range 1-9) and 2 (range 1-7), respectively, and there was a significant difference in the COI between the two infections (Wilcoxon test, $p < 0.0001$). Overall, both asymptomatic (72.9% vs. 27.1%) and febrile infections (53.1% vs. 46.9%,) had more samples with polyclonal infections compared to those with mono-infections (**Figure 3.2**).

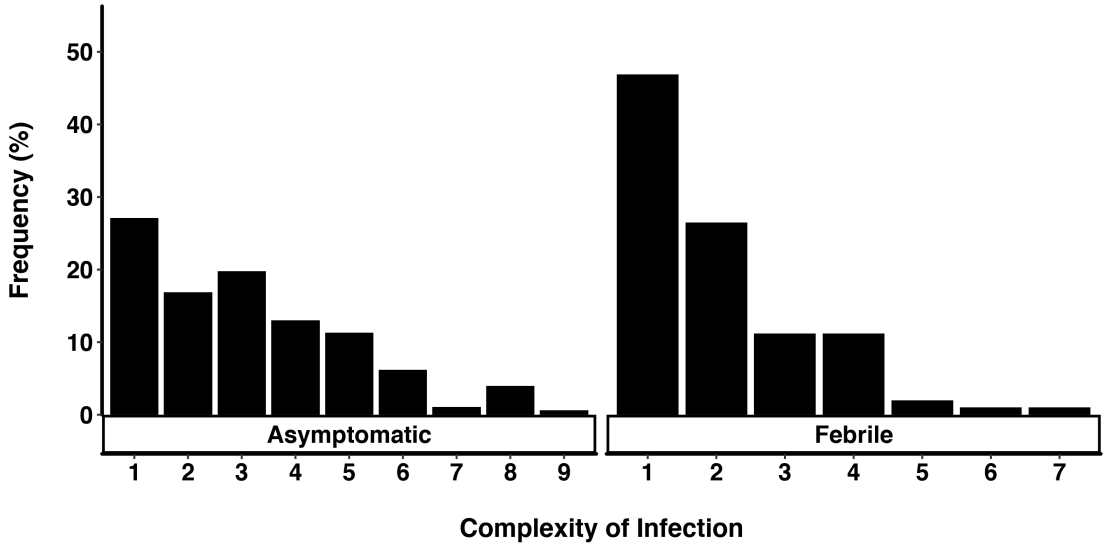


Figure 3.2. The distribution of complexity of infection (COI) between the asymptomatic and febrile malaria infections. Below the plot are the numbers 1-9 representing COI while on the right are the panel labels representing asymptomatic and febrile malaria samples. Above each bar is the percentage frequency of each of the COI in the population and no participant had a COI above 7 under the febrile malaria infections.

3.3.3 Temporal changes in *ama1* variant frequencies

Fewer samples were obtained from 2011-2014 for asymptomatics and from 2012-2016 for febrile malaria infections, hence, the data from these time-points were combined. In the asymptomatic infections, the most prevalent variant in 2007 (PFAMA1.00) was detected across all time-points with frequencies ranging between 6-27% and was the most dominant in three of the five time-points. No other variant was detected across all time-points, however, three variant were detected in four of the five time-points including PFAMA1.02 (range 3-8%, 2007-2010), PFAMA1.05 (range 3-9%, 2007-2009 and 2011-2014) and

PFAMA1.09 (range 4-7%, 2008, 2009, 2010 and 2011-2014) and two variants were detected in two of the five time-points including PFAMA1.01 (2008-2010) and PFAMA1.08 (2007, 2008 and 2010). All other variants were detected in less than three time-points each (**Table 3.10**).

In the febrile infections, the most dominant variant in 2007 (PFAMA1.01) was detected across all time-points with frequencies ranging between 6-27%, and was only dominant in 2007. Two variants were detected in three of the four time-points including PFAMA1.00 (range 5-9%, 2009, 2010 and 2012-2016), PFAMA1.05 (range 4-13%, 2009, 2010 and 2012-2016), PFAMA1.07 (range 3-4%, 2009, 2010 and 2012-2016) and PFAMA1.10 (range 4-10%, 2009, 2011 and 2012-2016). All other variants were detected in less than three time-points each (**Table 3.10**).

Table 3.10 *ama1* variant frequencies in asymptomatic and febrile infections over time

Variant ID	Asymptomatic Infections					First-Febrile Infections			
	2007	**2008**	**2009**	**2010**	**2011-2014**	**2009**	**2010**	**2011**	**2012-2016**
PFAMA1.00	29.25	16.45	11.23	6.54	19.12	5.51	7.29	2.63	9.29
PFAMA1.01	0.54	4.42	14.87	4.6	1.25	27.06	9.64	6.55	9.77
PFAMA1.02	3.69	4.54	6.48	8.17	-	18.91	10.85	-	0.06
PFAMA1.03	1.07	2.18	7.25	0.27	-	-	-	-	-
PFAMA1.04	-	-	-	1.42	-	-	-	-	-
PFAMA1.05	8.87	3.71	4.89	1.02	9.16	-	8.24	4.32	13.62
PFAMA1.06	4.79	9.18	-	5.29	-	-	0.41	-	0.17
PFAMA1.07	1.62	4.12	2.69	1.21	6.94	0.32	5.99	4.35	0.05
PFAMA1.08	0.87	4.2	4.41	4.95	7.14	0.22	1.37	6.93	-
PFAMA1.09	2.65	0.45	2.7	5.25	2.16	3.68	3.32	3.86	4.17
PFAMA1.10	4.33	1.08	0.27	1.22	7.14	5.72	2.37	10.95	4.43
PFAMA1.11	2.29	3.02	2.6	-	7.14	1.9	2.78	1.67	4.05
PFAMA1.12	-	3.72	2.48	4.32	-	-	0.88	-	2.61
PFAMA1.13	0.65	1.64	-	0.35	-	-	3.03	10.05	-
PFAMA1.14	2.9	1.14	-	1.64	0.04	-	0.12	-	0.83
PFAMA1.15	-	3.3	5.13	3.24	-	-	4.68	-	-
PFAMA1.16	-	1.57	-	1.74	1.61	-	2.59	3.27	2.76
PFAMA1.17	-	-	3.05	1.76	11.55	-	0.17	3.27	-
PFAMA1.18	-	1.31	8.27	0.94	-	-	0.39	-	2.53
PFAMA1.19	4.35	-	0.12	2.81	-	-	2.7	-	3.94
PFAMA1.20	-	1.14	-	1.38	7.14	-	1.59	-	2.98
PFAMA1.21	-	0.64	0.6	3.76	-	0.78	-	0.13	-
PFAMA1.22	-	2.25	-	0.16	-	2.68	2.78	0.18	1.64
PFAMA1.23	1.84	0.06	-	2.25	-	0.55	0.15	-	-
PFAMA1.24	1.66	2.91	-	0.83	4.66	-	-	-	-
PFAMA1.25	-	0.69	-	1.49	-	-	0.47	4.35	-
PFAMA1.26	-	0.22	1.75	0.82	-	-	0.08	0.42	0.15
PFAMA1.27	-	-	-	3.71	-	2.99	0.19	-	2.79
PFAMA1.28	-	0.23	-	1.91	-	4.46	2.19	-	-

PFAMA1.29	3.31	0.64	1.87	0.48	-	-	2.17	-	-
PFAMA1.30	-	-	-	-	0.57	0.16	0.36	4.42	4.23
PFAMA1.31	-	1.58	1.14	0.44	-	-	0.29	-	0.26
PFAMA1.32	6.8	0.21	-	-	7.14	-	-	0.03	-
PFAMA1.33	-	0.63	2.9	0.09	-	2.99	-	4.09	-
PFAMA1.34	-	3.88	-	-	-	-	-	0.48	2.69
PFAMA1.35	-	0.3	1.26	1.05	-	-	-	4.32	0.47
PFAMA1.36	-	1.08	0.26	1.44	-	-	0.07	-	-
PFAMA1.37	-	-	0.14	0.81	-	1.97	-	0.03	-
PFAMA1.38	8.76	-	-	-	-	-	-	-	-
PFAMA1.39	-	1.52	0.36	0.08	-	-	2.8	-	-
PFAMA1.40	-	0.13	3.15	0.1	-	-	-	-	-
PFAMA1.41	-	-	0.1	0.32	-	0.13	0.37	-	-
PFAMA1.43	-	-	-	2.61	-	5.56	-	-	-
PFAMA1.44	-	-	-	0.97	-	-	1.17	4.3	1.04
PFAMA1.45	-	1.32	-	0.99	-	-	-	-	4.17
PFAMA1.46	-	-	0.13	1.89	-	-	2.55	-	-
PFAMA1.47	-	-	-	1.43	0.07	-	-	-	5.15
PFAMA1.48	-	-	0.72	1.59	-	-	-	-	-
PFAMA1.49	-	-	-	-	-	-	3.3	-	-
PFAMA1.50	0.83	-	0.29	0.35	-	-	1.39	-	-
PFAMA1.51	-	2.18	-	-	-	-	0.02	-	0.02
PFAMA1.52	0.45	-	-	-	-	-	2.78	-	4.17
PFAMA1.53	-	0.28	-	2.34	-	-	-	-	-
PFAMA1.54	-	2.75	-	-	-	-	-	-	-
PFAMA1.55	0.45	-	-	0.12	-	-	2.78	-	-
PFAMA1.56	-	0.25	-	1.32	-	-	0.21	-	-
PFAMA1.57	-	-	3.59	0.07	-	-	-	-	-
PFAMA1.58	4.17	-	-	-	-	-	-	4.35	-
PFAMA1.59	-	-	-	0.66	7.14	-	-	-	-
PFAMA1.60	-	-	-	0.53	-	-	2.5	-	-
PFAMA1.61	-	-	-	1.43	-	-	-	-	-
PFAMA1.62	-	0.26	-	-	-	-	-	-	-
PFAMA1.63	-	-	-	1.54	-	-	-	-	-
PFAMA1.64	-	-	-	-	-	5.86	-	-	-
PFAMA1.65	-	2.14	-	-	-	-	0.21	-	-
PFAMA1.66	-	0.04	-	-	-	5.56	-	-	-
PFAMA1.67	-	-	-	-	-	-	0.03	4.35	-
PFAMA1.68	-	-	2.53	0.08	-	-	-	-	-
PFAMA1.69	-	1.67	-	-	-	-	-	-	-
PFAMA1.70	-	0.03	-	-	-	3	-	-	-
PFAMA1.71	-	1	-	0.1	-	-	-	-	-
PFAMA1.72	0.92	-	-	-	-	-	-	-	-
PFAMA1.73	-	-	-	-	-	-	2.78	-	-
PFAMA1.74	-	-	-	1.43	-	-	-	-	-
PFAMA1.75	-	-	-	-	-	-	-	4.35	-
PFAMA1.76	-	-	-	-	-	-	-	4.35	-
PFAMA1.77	-	2.24	-	-	-	-	-	-	-
PFAMA1.78	-	-	-	-	-	-	-	-	4
PFAMA1.79	-	-	-	-	-	-	-	-	3.98
PFAMA1.80	-	-	-	-	-	-	-	-	3.98
PFAMA1.81	-	-	2.79	-	-	-	-	-	-
PFAMA1.82	2.94	-	-	-	-	-	-	-	-

PFAMA1.83	-	-	-	0.96	-	-	-	-	-
PFAMA1.84	-	-	-	0.93	-	-	-	-	-
PFAMA1.85	-	-	-	-	-	-	-	2.01	-
PFAMA1.86	-	0.9	-	-	-	-	-	-	-
PFAMA1.87	-	0.79	-	-	-	-	-	-	-
PFAMA1.88	-	-	-	0.42	-	-	-	-	-
PFAMA1.89	-	-	-	0.39	-	-	-	-	-

The table shows the frequencies (percentage) of different *ama1* variants over time in both asymptomatic and febrile infections. Under "Variant ID" are the *ama1* variants numbered from 00 to 89. Below both asymptomatic and febrile infections are the sampling timepoints and marked with "-" are timepoints where the respective *ama1* variants were not detected.

3.3.4 Temporal changes in Complexity of Infection

A comparison of temporal changes in population COI in asymptomatic infections revealed that there was a significant drop in median COI from earlier time-points (2007-2010) (median COI = 3.5) to 2011-2014 (median COI = 1) (Kruskal-Wallis test *p-value* = 0.0013). Additionally, there was a significant drop in COI between 2009 and 2010 (COI = 3 to 2, Wilcoxon test, $p < 0.001$) and 2010 and 2011-2014 (COI = 2 to 1, Wilcoxon test, $p < 0.01$). As for the febrile malaria infections, there was no significant change in median COI over time (Kruskal-Wallis test *p-value* = 0.61) (**Figure 3.3**)

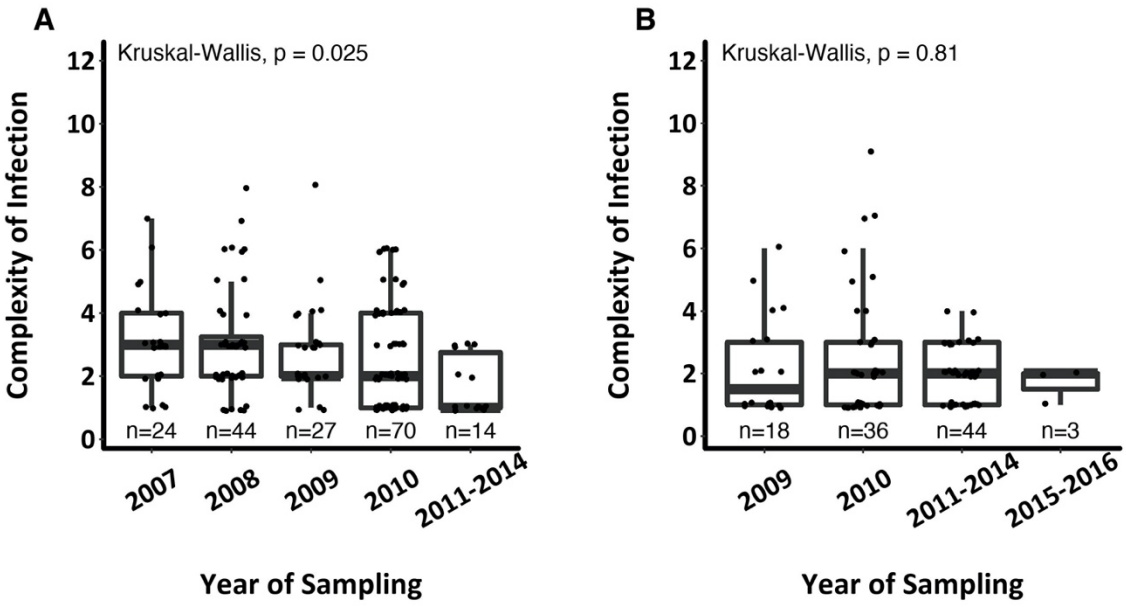


Figure 3.3. Temporal changes in median COI in *(A)* asymptomatic infections and *(B)* febrile infections. A significant temporal drop in COI was observed in asymptomatic infections, however, this drop was not observed in febrile malaria infections.

3.3.5 Transition from asymptomatic to first febrile infection

Of the 30/114 paired asymptomatic-febrile infection samples that were successfully sequenced, four febrile infections were as result of an *ama1* variant detected in the preceding asymptomatic infection. A half of these four infections involved a decrease in COI while the other half showed no change in COI and all four cases were from participants >5 years of age (**Table 3.11**). for all the four participants, the likelihood that the persistent variants were acquired from a new infectious mosquito bite was low.

The subsequent febrile malaria episode for Participant 6 contained two DNA sequence variants that were persistent. This included PFAMA1.01 that decreased from 95.3% in the prior asymptomatic infection to 43.3% and PFAMA1.00 that increased from 0.42 to 3.22% to become the second most and least dominant variants, respectively. Participant 07's subsequent febrile malaria episode harboured a persistent PFAMA1.07 that increased in frequency from 0.22% in the prior asymptomatic infection to 65.6%. Participant 13's subsequent febrile infection harboured a persistent PFAMA1.14 variant as the least dominant variant and its frequency increased from 0.9% in the prior asymptomatic infection to 5.8%.

finally, participant 46's subsequent febrile infection harboured two persistent variants PFAMA1.01 and PFAMA1.00 that increased from 47% to 64.4% and from 5.5% to 9.8% to become the most dominant and least dominant variants, respectively.

Table 3.11. Transition from asymptomatic to first febrile infection with a persistent clone

PID	Change in COI	Variant ID	Frequency		Characteristic	Binomial probability of reinfection
			Asymptomatic	Febrile		
PID_06	No Change in COI (4 to 4)	PFAMA1.01	95.4%	43.3%	Persistent	0.21
		PFAMA1.10	0%	8.9%	Novel	
		PFAMA1.31	0%	44.5%	Novel	
		PFAMA1.03	3.7%	0%	Cleared	
		PFAMA1.41	0.5%	0%	Cleared	
		PFAMA1.00	0.4%	3.2%	Persistent	
PID_07	Decreased COI (3 to 2)	PFAMA1.06	73.1%	0%	Cleared	0.06
		PFAMA1.03	26.7%	0%	Cleared	
		PFAMA1.07	0.2%	65.6%	Persistent	
		PFAMA1.02	0%	34.4%	Novel	
PID_13	Decreased COI (6 to 4)	PFAMA1.21	56.1%	0%	Cleared	0.1
		PFAMA1.18	30.4%	0%	Cleared	
		PFAMA1.17	6.9%	0%	Cleared	
		PFAMA1.10	5.1%	0%	Cleared	
		PFAMA1.14	0.9%	5.8%	Persistent	
		PFAMA1.02	0.7%	0%	Cleared	
		PFAMA1.48	0%	49.4%	Novel	
		PFAMA1.00	0%	34.5%	Novel	
		PFAMA1.15	0%	10.3%	Novel	
PID_46	No Change in COI (3 to 3)	PFAMA1.05	47.4%	0%	Cleared	0.27
		PFAMA1.00	47.1%	64.4%	Persistent	
		PFAMA1.01	5.5%	9.9%	Persistent	
		PFAMA1.53	0%	25.7%	Novel	

Under the column "characteristic", persistent represents a variant that was detected in both the asymptomatic and febrile malaria infections, "novel" represents a variant that was detected in the febrile malaria infection only, while "cleared" represents a variant that was only detected in the prior asymptomatic infection only. The grey cells represent points where the respective variants were not detected. The binomial probability of reinfection represents the likelihood that the persistent variant was acquired from a subsequent mosquito bite (reinfection). Entries marked with "-" indicate variants that were not detected. PID – Participant ID.

26 febrile malaria infections were as a result of an *ama1* variant that was not detected in the asymptomatic infection. Four (15.3%) of the transitions did not show a change in COI and one was from a participant of ≤5 years of age, while the others were from participants >5 years of age (**Table 3.12A**). Eleven of the transitions (42.3%) involved a decrease in COI and all the participants were > 5 years of age (**Table 3.12B**). The final eleven transitions (42.3%) involved

an increase in COI with four participants being ≤5 years of age while 7 were >5 years of age (**Table 3.12C**).

For the four participants that developed a febrile malaria episode with no change in COI, none had a COI above three as their COI ranged between 1 and 2. For the eleven participants that developed a febrile malaria episode with a decrease in COI, ten participants had asymptomatic infections with a COI greater than two (range 3-7) while one participant had a COI of two. Ten of these participants then went on to develop a febrile episode with a COI of one, while one developed a febrile episode with a COI of two. For the eleven participants that developed a febrile malaria episode with an increase in COI, all participants had an asymptomatic infection with a COI of two and below. All these eleven participants went on to develop a febrile malaria episode with a COI of two and above (range 2-4)

Table 3.12. Transition from asymptomatic to first febrile infection with a novel clone

A

PID	COI Change	Variant ID	Frequency	
			Asymptomatic	Febrile
PID_20	No Change in COI (1 to 1)	13	100%	0%
		6	0%	100%
PID_37	No Change in COI (2 to 2)	6	98.2%	0%
		0	1.8%	0%
		20	0%	94.7%
		10	0%	5.3%
PID_44	No Change in COI (1 to 1)	14	100%	0%
		5	0%	100%
PID_45	No Change in COI (1 to 1)	11	100%	0%
		8	0%	100%

B

PID	COI Change	Variant ID	Frequency	
			Asymptomatic	Febrile
PID_01	Decreased COI (4 to 2)	28	54.9%	0%
		9	36.2%	0%
		2	4.8%	0%
		50	4.0%	0%
		1	0%	100%
PID_02	Decreased COI (7 to 1)	0	27.0%	0%
		6	61.4%	0%
		7	7.0%	0%
		3	3.0%	0%
		5	1.0%	0%
		40	0.3%	0%
		12	0.3%	0%
		1	0%	100%
PID_11	Decreased COI (5 to 1)	5	98.7%	0%
		42	0.5%	0%
		0	0.3%	0%
		22	0.2%	0%
		26	0.2%	0%
		64	0%	100%
PID_14	Decreased COI (3 to 1)	4	89.0%	0%
		3	10.2%	0%
		39	0.8%	0%
		5	0%	100%
PID_15	Decreased COI (3 to 1)	39	77.8%	0%
		34	21.9%	0%
		4	0.3%	0%
		59	0%	100%
PID_19	Decreased COI (5 to 1)	30	29.8%	0%
		27	25.9%	0%
		2	18.6%	0%
		14	15.8%	0%
		41	9.9%	0%
		17	0%	100%
PID_23		16	44.2%	0%

C

PID	COI Change	Variant ID	Frequency	
			Asymptomatic	Febrile
PID_05	Increased COI (2 to 4)	18	99.3%	0%
		5	0.7%	0%
		10	0%	93.6%
		33	0%	2.9%
		52	0%	2.3%
		7	0%	1.2%
PID_16	Increased COI (1 to 2)	36	100%	0%
		1	0%	88.7%
		2	0%	11.3%
PID_18	Increased COI (1 to 3)	27	100%	0%
		2	0%	90.2%
		6	0%	6.0%
		47	0%	3.9%
PID_21	Increased COI (1 to 4)	8	100%	0%
		30	0%	78.7%
		2	0%	9.4%
		27	0%	6.9%
		6	0%	5.0%
PID_22	Increased COI (1 to 2)	2	100%	0%
		20	0%	97.1%
		29	0%	2.9%
PID_24	Increased COI (1 to 3)	63	100%	0%
		47	0%	94.2%
		52	0%	4.9%
		0	0%	0.9%
PID_27	Increased COI (1 to 3)	35	100%	0%
		10	0%	51.7%
		11	0%	37.8%
		29	0%	10.5%
PID_33	Increased COI (1 to 4)	0	100%	0%
		1	0%	41.4%
		17	0%	35.1%
		9	0%	19.7%
		25	0%	3.8%

	Decreased COI (5 to 1)	7	33.3%	0%
		1	9.9%	0%
		25	6.6%	0%
		12	5.9%	0%
		13	0%	100%
PID_25	Decreased COI (6 to 1)	42	49.8%	0%
		12	16.8%	0%
		8	9.6%	0%
		0	9.0%	0%
		1	8.5%	0%
		2	6.3%	0%
		11	0%	100%
PID_31	Decreased COI (3 to 2)	14	63.4%	0%
		7	29.4%	0%
		33	7.2%	0%
		45	0%	98.8%
		1	0%	1.2%
PID_32	Decreased COI (2 to 1)	5	82.2%	0%
		0	17.8%	0%
		36	0%	100%
PID_36	Decreased COI (3 to 1)	21	65.3%	0%
		15	22.7%	0%
		1	12.0%	0%
		59	0%	100%
PID_35	Increased COI (1 to 3)	60	100%	0%
		15	0%	57.0%
		45	0%	21.7%
		32	0%	21.3%
PID_39	Increased COI (1 to 2)	10	100%	0%
		37	0%	12.2%
		11	0%	87.8%
PID_42	Increased COI (1 to 2)	19	100%	0%
		17	0%	98.5%
		33	0%	1.5%

The table has been three groups that included participants who developed a febrile episode with *(A)* no change in COI, *(B)* decrease in COI and *(C)* increase in COI compared to the prior asymptomatic infection. The grey cells represent points where the respective variants were not detected. PID – Participant ID.

Median time to infection in individuals who developed a febrile episode with a persistent clone versus a febrile episode with a novel clone was 66.5 days and 61.5 days, respectively. A statistical test could however not be applied due to the limited power (4 vs 26, respectively).

For asymptomatic infections, an increase in age was associated with a 3% increase in COI, however this was not significant (IRR = 1.03, 95% C.I = 0.99-1.07, p = 0.14), however, a log increase in parasitemia was significantly associated with a 4% increase in COI (IRR = 1.04, 95% C.I = 1.01-1.07, p = 0.008). For febrile malaria infections, an increase in age was associated with a decrease in COI, however this was not significant (IRR = 0.95, 95% C.I = 0.88-1.03, p = 0.2) and a log increase in parasitemia was associated with an increase in COI but this was also not significant (IRR = 1.03, C.I = 0.96-1.11, 0.4). Finally, a unit increase in COI was significantly associated with a 4% increase in time to first febrile malaria episode (IRR = 1.04, 95% C.I = 1.04-1.05, p < 0.001).

3.4 Discussion

In this chapter the major findings include the observation that both asymptomatic and febrile malaria infections harboured genetically diverse parasites with asymptomatic infections having higher COI compared to febrile malaria infections. The prevalence of variants detected changed drastically from one year to the next. Thus, some variants were observed once throughout the entire sampling timeframe, while others appeared and disappeared, only to reappear later and this was consistent even with the rare variants. Both the asymptomatic and febrile malaria infections shared *ama1* variants, however, the most dominant *ama1* variants in both infections were distinct. Also, a temporal trend in the decrease of COI in asymptomatic infections was observed, however, this was not evident for febrile malaria infections. Few individuals maintained a persistent clone between asymptomatic and febrile infection pairs; hence, a large majority of the febrile infections were associated with the introduction of new clones. Remarkably, the median time to infection was longer in the individuals who developed a febrile malaria episode with a persistent clone, albeit this finding was made in a small sample size. Furthermore, it was not possible to ascertain whether the persistent clone was responsible for the subsequent febrile episode. An increase in

asymptomatic parasitemia at baseline was significantly associated with an increase in COI and an increase in asymptomatic COI at baseline was significantly associated with an increase in time to first febrile malaria infection. Finally, *ama1* has been reported to be one of the most polymorphic genes in *P. falciparum*. This was further shown when a comparing *ama1* haplotypes from the current study with those from other regions and little overlap was detected even over just a 459bp region.

This study is in agreement with previous work that showed that asymptomatic infections harbour more diverse parasites compared to febrile malaria infections (Beck *et al.* 1997; Magesa *et al.* 2002). Previous work has shown that asymptomatic individuals with a high COI have a broader spectrum of immunological memory compared to those with lower COI (Rono *et al.* 2013). Whether the presence of multiclonal infections leads to increased breadth, or whether breadth facilitates tolerance of multiclonal infections, is difficult to resolve. On the other hand, lower COI in febrile malaria episodes could be explained by the fact that most fever episodes appear to be associated with parasitaemia peaks (Cox *et al.* 1994) and this may result from the expansion of just a few or even single genotypes. Therefore, this could lead to sampling of only a few clones leading to minority clones remaining undetectable.

Regarding temporal changes in variant frequencies, previous studies based on *msp1*/*msp2* genotyping (Daubersies *et al.* 1996; Babiker *et al.* 1998) and SNP genotyping (Nkhoma *et al.* 2018), have revealed that asymptomatic infections exhibit a rapid turnover of clones over time. Daubersies et al. (1996) sampled individuals at two-week intervals over a period of three months and showed that alleles appeared and disappeared through time. Many alleles were detected over a period of 2-3 weeks, but some alleles could be detected only for a few days. Babiker et al. (1998) sampled individuals every two weeks for two months and subsequently once each month for a further 15 months and showed that single genotypes within multiclonal infections fluctuated considerably during this period. Finally, Nkhoma et al. (2018) sampled individuals over seven consecutive days and showed that even for a shorter time frame, the genetic diversity within individuals changed dramatically over the course of follow up. In this current study, even though sampling was conducted from one year to another and

with varying sample sizes, overall, there was a rapid turnover of *ama1* variants highlighting how heterogenous the parasite population is in asymptomatic infections. Nevertheless, using shorter sampling timeframes, future studies should attempt to characterise multiclonal infections using amplicon-deep sequencing to reveal the full extent of genetic diversity in such infections.

Concerning the differences in predominant variants in asymptomatic and febrile malaria infections, previous studies based on *msp1*/*msp2* genotyping have also reported differences in carriage of the different *msp1* (K1, MAD20 and RO33) and *msp2* (3D7 and FC27) alleles. Based on *msp1*-genotyping some studies have reported that the K1 allele predominates in febrile malaria infections (Somé *et al.* 2018; Sondo *et al.* 2019) while some have found the MAD20 allele do be the most dominant (Chen *et al.* 2018). Based on *msp2*-genotyping, previous studies have reported that the *msp2*-3D7 is predominant in asymptomatic infections (Bereczky *et al.* 2007; Färnert *et al.* 2009) while others have reported that the FC27 allele predominates (Sondén *et al.* 2015). As for febrile malaria infections, certain studies have reported that the 3D7 allele predominates (Bereczky *et al.* 2007; Mohammed *et al.* 2015; Somé *et al.* 2018; Sondo *et al.* 2019) while others have reported that the FC27 allele predominates (Chen *et al.* 2018). While the current study could not be compared to previous work since it was based on *ama1* SNP genotyping, it does show that asymptomatic infections tend to be caused by clones that are predominantly distinct from those that are responsible for febrile malaria episodes and further work is needed to understand this.

Previous work has found COI to be a surrogate marker for malaria transmission intensity, where multiple clone infections are expected to occur more frequently when transmission levels are higher. Using *msp1*/*msp2* genotyping, Bendixen et al. (2001) showed that COI decreased from 3.5 to 1.6 when comparing villages in Tanzania in high and low malaria transmission intensities. Using high-throughput SNP genotyping, Nkhoma et al. (2013) showed that over a 10 year period (2001-2010), reduced malaria transmission was accompanied by a significant decrease in the frequency of multi-clonal infections (63% to 14%) at the Thai-Burma border. Finally, using a 24 SNP barcode, a recent study conducted in Senegal showed that COI decreased from 1.35 to 1 when comparing samples originating from

a period of relatively high malaria transmission and samples collected during a period of relatively low malaria transmission, respectively (Bei *et al.* 2018). In agreement with these studies, the current study also showed decreases in COI in tandem with decreasing malaria transmission intensity highlighting the utility of using COI data to measure changes in *P. falciparum* transmission.

Previous studies have reported the introduction of novel variants in febrile episodes following an asymptomatic infection, suggesting that the parasites causing a clinical episode are those against which an individual has not yet mounted an efficient protective immune response (Roper *et al.* 1998; Ofosu-Okyere *et al.* 2001; Kun *et al.* 2002; Buchwald *et al.* 2018). This is in agreement with the current study. Furthermore, previous studies have shown that after an asymptomatic infection, subsequent clinical malaria is associated with the same clones in regions of low transmission (Roper *et al.* 1998; Nsobya *et al.* 2004) and with an increase in COI in regions of high transmission (Ofosu-Okyere *et al.* 2001; Mueller *et al.* 2012). In contrast, no study was identified that showed a decrease in COI during this transition. The current study was conducted in a region of moderate to high malaria transmission intensity and in line with previous studies, it was expected that the transition from asymptomatic to febrile malaria will mainly be associated with an increase in COI. Conversely, patterns of no change in COI, a decrease and increase in COI were observed. Malaria transmission in the area under study has been shown to be heterogenous (Bejon *et al.* 2010) and the individuals who showed no change in COI in the subsequent febrile malaria infection could be living in homesteads where malaria transmission is generally lower compared to the overall malaria transmission intensity in the area. Indeed, while these participants were of >5 years of age, they had lower COI than their counterparts overall. Likewise, living in homesteads where malaria transmission is considerably less may have resulted in these participants acquiring less diverse subsequent febrile infections. For the participants that developed subsequent febrile malaria infection with an increase in COI, this could have been a result of a super infection after repeated bites from infected mosquitoes with genetically distinct parasites. For the participants that developed subsequent febrile malaria infection with a decrease in COI, it may be as a result of the participants having varied COI later on during their asymptomatic

infections before developing a febrile malaria episode or getting an infection from mosquitoes with less COI compared to the prior asymptomatic infection. The impact of age on the transition from an asymptomatic to febrile infection could not be tested as there was an over-representation of samples in the >5 years age group (25 vs. 5). Future studies can mitigate this by ensuring that they have a larger sample size across different age groups.

In previous prospective studies, asymptomatic COI detected at baseline has previously been shown to have an impact on the risk of developing febrile malaria. Accordingly, increasing COI has been associated with a reduced risk of subsequent febrile episodes (Al-Yaman *et al.* 1997; Färnert *et al.* 1999; Bereczky *et al.* 2007; Pinkevych *et al.* 2014; Sondén *et al.* 2015). However, some studies have reported the inverse, notably in children <5 years of age (Henning *et al.* 2004). The current study, showed that increasing COI was associated with an increase in time to first febrile episode. However, it is worth noting that the current study was not prospective and hence missed out on participants who did not develop a febrile episode and this may have biased the findings. Nevertheless, higher COI could be an indicator of increased acquired immunity such that younger children (with less effective NAI) are unable to tolerate polyclonal infections while older children can tolerate polyclonal infections and develop symptoms less frequently.

Finally, previous studies have reported increases in parasite density to be associated with increasing COI (Kateera *et al.* 2016; Nabet *et al.* 2016), similar to the current study. One hypothesis to explain this is that sampling greater number of parasites increases the likelihood of detecting mixed infections.

This study had several limitations, *(i)* the sequencing controls enabled the detection of variants with >5% frequency only, in each sample. However, if a variant was identified at <5% frequency in one sample but at >5% in another sample, this variant was retained as this increased the confidence that it may be a “true” variant. It is possible that minor variants (<5% frequency) might exist in smaller relative proportions so as to evade immune responses. Consequently, it’s worth including multiple sampling timepoints to increase the likelihood of capturing minor variants *(ii)* Many samples had low parasitaemia and this resulted in fewer samples being amplified and sequenced. Consequently, more than half of the total

asymptomatic samples were not sequenced; hence future studies should attempt to run nested PCR reactions to try and increase the PCR yield for sequencing samples with low parasitaemia. Accordingly, an earlier study by Takala et al. (2009) employed nested PCR to amplify the *ama1* gene in 506 *falciparum* infections from 100 individuals and identified 214 unique *ama1* variants. While Takala et al. (2009) amplified a larger *ama1* fragment (445-1656bp) that could have resulted in the detection of many more polymorphic loci and hence many more variants, it underscores the utility of nested PCR in increasing the PCR yield for sequencing. *(iii)* Sampling was conducted at only one time-point and previous studies have shown that a single peripheral blood sample, poorly tracks overall parasite diversity and should be considered as only snapshots in time (Daubersies *et al.* 1996; Babiker *et al.* 1998; Nkhoma *et al.* 2018). This may have limited the ability to describe the full extent of parasite diversity in asymptomatic infections. *(iv)* Sequencing was conducted in two separate labs and this may have introduced batch effects. Additionally, the KWTRP batch had higher sequencing depth compared to the IDEEL lab dataset. However, a majority (66.1%) of all the variants detected, overlapped between the two labs. For the variants that did not overlap, it is expected given the highly polymorphic nature of *ama1* and this increases the chances that two samples will present with different variants. *(v)* The current study relied on only one polymorphic marker, *ama1*, and a recent study demonstrated that the inclusion of several markers can help to compare diversity from multiple targets (Lerch *et al.* 2017). Future studies should aim to include more targets to reliably describe and compare parasite diversity, however, the use of only one highly polymorphic is appealing as including many markers might be resource intensive for use as a surveillance tool. *(vi)* In this region of moderate to high transmission, it is likely that some persistent variants were acquired from new mosquito infections and hence falsely classified. However, we found the probability of reinfection to be low. That said, there is still a chance that some of the persistent variants may have been acquired from new infections since the median time to reporting a first febrile episode was more than 30 days, and this would have coincided with the peaking of the malaria transmission season. *(vii)* Lastly, half the sequence reads were discarded during the analysis as their MID tags could not be identified. The forward primers were designed to include MID

tags and it is possible that a huge proportion of primers were unsuccessfully tagged, hence, any amplicons that were generated from these untagged primers would be discarded during sequence analysis. To avoid losing sequence data, future studies should try to quantify how much of their primers are actually tagged for example by sequencing the primer pool

Taken together, the current study found that asymptomatic infections harbour more diverse parasite populations compared to febrile malaria infections. Additionally, higher COI in asymptomatic infections predict the risk of subsequent febrile malaria episodes, suggesting that host immunity plays a prominent role in mediating this process. This finding is particularly interesting as it adds to the literature already showing that host immunity impacts on the risk of developing febrile malaria episodes. Consequently, the identification of the correlates of immunity will inform the development of a malaria vaccine that will aid in malaria elimination efforts.

Chapter 4 : Surveillance of *P. falciparum* drug-resistance markers over two decades of changing anti-malarial policies

4.1 Introduction

In addition to chronic asymptomatic infections, drug-resistance presents another obstacle to malaria elimination and this chapter focussed on carrying out surveillance of antimalarial drug-resistance in Kilifi, Kenya. The availability of longitudinal samples collected from the Kilifi County hospital from 1995-2016 prompted this study into carrying out a surveillance of antimalarial drug-resistance markers. Within this period, antimalarial drug policies in Kenya changed from using CQ, followed by SP and later to the introduction of ACTs due to widespread resistance to CQ and SP. Such data are important for assessing the efficacy of antimalarial drugs and can help to detect early signs of the emergence of drug-resistance.

There exist three main methods for assessing antimalarial drug-resistance and these include therapeutic efficacy studies (TES), *in vitro* and *ex vivo* phenotypic assays and finally the use of molecular markers of drug-resistance. TES is the gold standard used to guide treatment policy in malaria-endemic countries and these involve the prescription of the required doses of anti-malarial drugs to patients with *Plasmodium* parasites, presenting with uncomplicated malaria. The patients are then followed up by parasitological and clinical assessments from 28 to 63 days, depending upon the half-life of the medicine assessed, after which the treatment outcome is determined as successful or not (WHO 2009). In instances of clinical or parasitological failure, molecular genotyping is recommended to distinguish recrudescence from re-infection, using three highly polymorphic genes (*msp1*, *msp2* and *glurp*) and the results from genotyping are routinely detected by agarose gel electrophoresis. Depending on the quality of the gel and technical expertise, there can be variation in the interpretation of gel electrophoresis leading to inaccurate estimates, especially in areas of high transmission intensity due to the high number of complex infections (Mugittu *et al.* 2006; Greenhouse *et al.* 2007).

In vitro and *ex vivo* phenotypic assays test for anti-malarial susceptibility in parasites and can be performed on parasites collected from patients (*ex vivo*) or with culture-adapted parasites (*in vitro*). Parasites are cultured in the presence of anti-malarial drugs at varying concentrations to determine the growth inhibitory effect of a drug, or by exposing them to a specific high concentration of drug for a relatively short period (Witkowski *et al.*

2013). Parasite growth is then measured using various techniques, including microscopy, enzyme-linked immunosorbent assays (ELISA), SYBR green fluorescence and flow cytometry. These results are then used to determine either the concentration that inhibits parasite growth by 50% (50% inhibitory concentration - IC_{50}) (Woodrow *et al.* 2013) or the parasites' survival rate (Witkowski *et al.* 2013). The major drawback with using *in vitro* and *ex vivo* assays is that only a limited number of laboratories in malaria-endemic countries have access to the sophisticated equipment, extensive resources, training and expertise required.

The use of molecular markers of drug resistance involves the detection of genetic polymorphisms associated with anti-malarial drug resistance. This was made possible through the association of antimalarial drug-resistance with single nucleotide polymorphisms (SNPs) and amplification of genes coding for drug target proteins or transporters. Consequently, these have allowed for a better understanding of the emergence and spread of antimalarial drug resistance (Picot *et al.* 2009). This technique involves DNA extraction, followed by the amplification of the genes or loci of interest using PCR and later sequencing to detect SNPs or amplifications. Molecular markers have several advantages including the possibility of studying many isolates within a short time for example with DNA from dried blood spots (DBS) that are easy to collect, transport and store. Also, sample pooling strategies, advances in PCR and DNA sequencing technologies have improved the scalability of genotyping resistance markers (Taylor *et al.* 2013).

The emergence of artemisinin and partner drug-resistant *P. falciparum* in Cambodia and its spread to neighbouring countries in SE Asia has added on to the growing evidence regarding the utility of assessing molecular markers of resistance (Hamilton *et al.* 2019). Consequently, regions outside SE Asia are encouraged to continue surveillance of drug-resistance markers to enable the early detection of ACT drug-resistance. One of the ways that have been employed to track drug-resistance markers is through genetic epidemiology of drug-resistance markers over space and time. In turn, this enables the evaluation of the pattern of evolution of genes conferring resistance to antimalarials to detect parasite adaptation due to changing antimalarial drug policy. In Kenya, the genetic epidemiology of *crt*, *mdr1*, *dhfr* (markers of chloroquine [CQ] and sulfadoxine-pyrimethamine – [SP] resistance) revealed that

since the withdrawal of CQ and SP as first-line antimalarial treatments, chloroquine resistance (CQR) markers have reduced in frequency in the population while those of SP resistance have remained high (Okombo *et al.* 2014; Achieng *et al.* 2015). This can also be replicated to assess whether similar signatures have occurred with artemisinin resistance markers in Kilifi, in response to the adoption of ACTs as first-line antimalarial drugs in Kenya since 2004.

crt, *mdr1 and dhfr* have well described selection patterns in response to the withdrawal of CQ and SP. Notably, there is good evidence of selection by ACT use in Kenya towards a predominance of wild-type alleles in *crt* and *mdr1* and towards the fixation of the mutant *dhfr* alleles in a previous study conducted in Kilifi (Okombo *et al.* 2014). Additionally, evidence of ACT selection upon the *ap2-mu* and *ubp-1* locus was found in a study conducted in Western Kenya (Henriques *et al.* 2014) and this has been strengthened by recent *in vitro* studies of CRISPR-Cas9-generated variants of *ap2-mu* and *ubp-1*, which both elicit significantly reduced artemisinin susceptibility *in vitro* (Henrici, van Schalkwyk and Sutherland 2019). Likewise, *in vitro* studies using a *P. falciparum* clone with a SSA genetic background reported selection in the *falcipain-2a* gene after artemisinin selection pressure (Ariey *et al.* 2014). More recently, a study conducted in The Gambia also reported selection in a novel marker, *nfs*, associated with lumefantrine tolerance that requires confirmation in other geographical settings (Amambua-Ngwa *et al.* 2018). Therefore, the aim of this study was to use the introduction of ACT as the pivotal point to test for evidence of selection in novel antimalarial drug resistance markers (*ap2-mu*, *falcipain-2a*, *k13*, *nfs* and *ubp-1*) as well as artemisinin resistance predisposing mutations (*arps10* codon V127M*, crt* codon I356T*, fd* codon D193Y and *mdr2* codon T484I), whose impact are not yet fully understood.

4.2 Methods

4.2.1 Study Design

Parasite DNA was extracted from frozen blood samples obtained from patients presenting to Kilifi County Hospital with uncomplicated malaria (1 month – 15 years of age) in 1995/1996, 1999/2000, 2006/2007 and 2012/2013 from a previous study (Okombo *et al.* 2014) and additional samples from 2015/16 and 2017/18 from this new study. A hundred and fifty

samples were collected for each time-point and the samples span 24 years of changing drug policy, except for 2017/18 in which 109 samples were used (**Figure 4.1**).

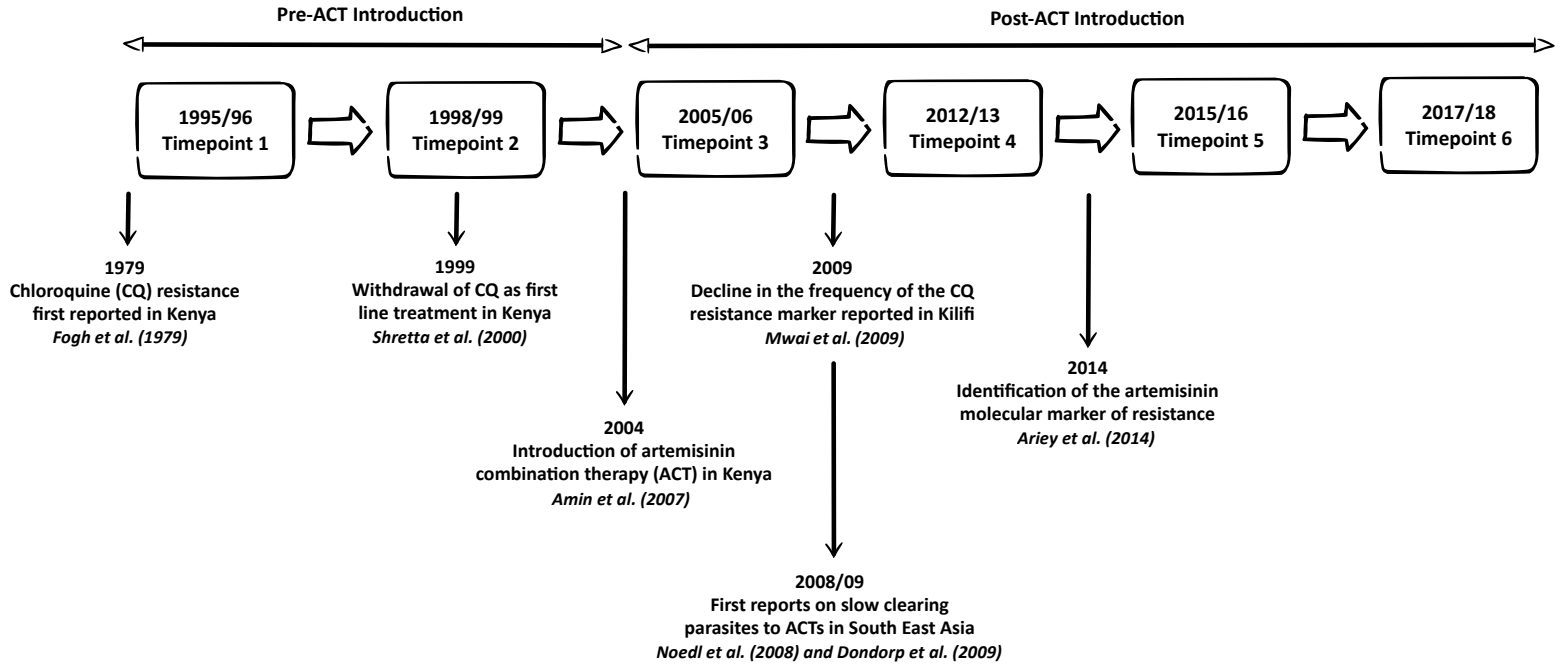


Figure 4.1. Schematic showing the time-points from which parasite populations were genotyped. Also indicated are historical highlights of antimalarial drug-resistance.

4.2.2 DNA extraction and PCR amplification

P. falciparum genomic DNA was extracted from frozen blood using the Qiagen DNA Blood Mini Kit (Qiagen, UK) as per the manufacturer's instructions (*see section 3.2.2*). Amplicons were generated from the following genes using primers from previous studies and primers designed in this study (**Table 4.1**): *crt* (PF3D7_0709000), *mdr1* (PF3D7_0523000), *dhps* (PF3D7_0810800), *nfs* (PF3D7_0727200), *k13* (PF3D7_1347700), *ap2mu* (PF3D7_1218300), *falcipain-2a* (PF3D7_1347700), *ubp-1* (PF3D7_0104300) and *serine-tRNA ligase, putative* (PF3D7_0717700). Additionally, four artemisinin resistance predisposing mutations; *arps10* (PF3D7_1460900), *crt* (PF3D7_0709000), *fd* (PF3D7_1318100) and *mdr2* (PF3D7_1447900) were genotyped. For PCR, the Expand™ High Fidelity PCR System, 0.5μl of template DNA, primers and the conditions indicated in **Table 4.1** were used. The final reaction volume was 10μl and the PCR amplification products were visualised on 1% agarose gels stained with

RedSafe™ Nucleic Acid Staining Solution (iNtRON Biotechnology DR). PCR negative samples were taken through a second and final round of PCR with 0.75μl of template DNA.

Table 4.1. List of primers for PCR and sequencing

Gene	Primer ID	Primer Sequence in '5-3' Orientation	PCR Annealing Temperature	Reference
ap2-mu (full length)	ap2mu-1F*	GTT AAC ACG ATT AGC GTC ATT TG	54°C for 2 min	Henriques et al., (2014)
	ap2mu-2R	GTC CTA TTA TGT ATA TGT GGA TC		
	ap2mu-3F	GAT ATC CAC AAA CAT TAG AAG TG		
	ap2mu-4R	CCA TCT GGT GGT GTG AAG G		
	ap2mu-5F	GCA TAT TTC ATC ATT GTG TTA CC		
	ap2mu-6R*	ACA CCC GAT TGA ACT ATT TAT AC		
crt (codon 72-76)	F*	GGT GGA GGT TCT TGTCTT GG	52°C for 30 sec	Okombo et al., (2014)
	R*	ATA AAG TTG TGA GTT TCG GAT G		
dhps (codons A437G and K540E and A581G)	F*	CCT AAA CGT GCT GTT CAA AGA ATG	58°C for 30 sec	Designed in this study
	R*	CAT CCA ATT GTG TGA TTT GTC CAC		
falcipain-2a (full length)	fp2F*	TGT AGC AAG AAC GTT TTG TGT AAA T	56°C for 2 min	Conrad et al., (2014)
	fp2NF	TGT GTA AAT TAA AGA TAA AAG TGC AAA		
	fp2R*	GGT AAA GGA AAA ATT AGT AAG GAT GC		
	fp2intR	GCA TAT TGT GAT TCT ACG GAA CC		
	fp2intF	AAA AAG CCC TAA TGG CAA GAA		
	fp2NR	GGT CCC TTT TTA AAA TAC TAT TGA CA		
k13 (full length)	k13-F*	ATG GAA GGA GAA AAA GTA AAA AC	56°C for 30 sec	Designed in this study
	k13-F2*	GCA GCA AAT CTT ATA AAT GAT G		
	k13-F3	GAA GCC TTG TTG AAA GAA GC		
	k13-F4	CAT AGG AAA CGA TTT GATG		
	k13-F1	CAA AAG CAA ATA GTA TCT CG		
	k13-R	CTC TTT TTT GTT GGT ATT CAT AAT TG		
	k13-R1*	CAC TAG CAT CAC TTA ATT CCG		
	k13-R2	CAC ATA CGC CAG CAT TG		
	k13-R3*	CGG AGT GAC CAA ATC TGG		
	k13-R4	CAT AGG AAA CGA TTT GAT G		
arps10 (codon V127M)	arps-F*	CAC AAT ATT ATG TTT CAT TTT AG	55°C for 30 sec	Designed in this study
	arps-R*	GTA TAA TTT ATT CTG CTT ACA TTC		
crt (codon I356T)	crt-F*	GAT TAT CGA CAA ATT TTC TAC	55°C for 30 sec	
	crt-R*	CTT TTT AAT TCT TAC GGC TAA G		
fd (codon *D193Y)*	fd-F*	GAT GCT AGT GAA AGA CAG AAT G	55°C for 30 sec	
	fd-R*	CAC ATA TTT TTG ATT GAG GAC		
mdr2 (codon T484T)	mdr2r-F*	GAG GTT TGT GGT GTA TTA TTT TC	55°C for 30 sec	
	mdr2r-R*	GTT AAA CCT ATA AAT AAT ACA C		
mdr1 (codons N86Y, Y184F and D1246Y)	mdr1_86/184_F*	TCC CAT TAA AGC CTC TTC TA	51°C for 30 sec	Okombo et al., (2014)
	mdr1_86/184_R*	ATG GGT AAA GAG CAG AAA G		
	mdr1_1246_F*	ATT GAT GTA AGA GAT GAT GGT		
	mdr1_1246_R*	TAT TCC ATC TTG TGC TGA TAA		
nfs	F1*	CTT CAA TTT TGT AAT GAA ATT TCT TC	51°C for 30 sec	Designed in this study
	R1*	CAT ATT ATG TCC ATA TAA ACT TTG GA		
	F2	ATG CAC GAA CAA ATC TTT TGA		
	R2	AGA TTT GTT CGT GCA TCC TCC GA		
ubp-1	F*	CGC CCG TAC TAT GAA GAA GAT C	50°C for 2 min	Henriques et al., (2014)

	R*	CCG TTT TAC CTG AAC TGT TCA GG		
serine-tRNA ligase, putative	RNAligase_F1*	ATG GTT TTA GAT ATA AAT TTA TTT CG	55°C for 30 sec	Designed in this study
	RNAligase_F2	AGA AAA TAG GTG GAG CTA		
	RNAligase_F3	ATG ATA CAA ACA TGT GAA GA		
	RNAligase_R1	TCT TGT TCA CTC CAT AAA GGG A		
	RNAligase_R2*	GTA ATA AGA AAC CTG CAC CTG		

All primers were used for Sanger sequencing while those marked with * were used for both PCR and sequencing. The Taq polymerase used in this study (Expand™ High Fidelity PCR system) has two rounds of PCR including a first-round that was set at 10 cycles and the second set at 25 cycles. The following conditions were used including the annealing temperatures indicated above: denaturation (94°C – 15 sec), annealing (gene-specific), extension (72°C – 2 min), and final extension (72°C – 7 min).

4.2.3 Sanger (chain-termination) sequencing

Positive PCR products were purified using ExoSAP-IT® (Thermo Fisher Scientific) and directly sequenced using the primers indicated in **Table 4.1** and the BigDye Terminator v3.1 Cycle Sequencing Kit v3.1 (Applied Biosystems, UK). Following PCR amplification, positive PCR products were purified using ExoSAP-IT™ Express (Thermo Fisher Scientific). This involved adding 1μl of ExoSAP-IT™ Express reagent into 5μl of the PCR product. This mixture was then incubated in a thermal cycler at 37°C for 4 min to degrade excess primers as well as nucleotides and another incubation at 80°C for 1 minute (min) to inactivate the ExoSAP-IT™ Express reagent. The PCR products were taken through a chain-termination reaction, as detailed below to sequence the DNA fragments.

The chain termination reaction was conducted using the BIG DYE terminator chemistry v3.1 kit (Applied Biosystems, UK) according to the manufacturer's instructions. Sequencing was done at the International Livestock Research Institute (ILRI, Kenya) using an ABI 3730xl capillary sequencer (Applied Biosystems, UK). Briefly, the PCR product was combined in a tube with a primer, DNA polymerase, deoxynucleotide triphosphates (dNTPs) and four fluorescently-labelled chain-terminating di-deoxynucleotide triphosphates (ddNTPs). ddNTPs are dNTPs that lack a 3'-hydroxyl group and this restricts the further extension of a copied DNA chain. The mixture was then transferred to a thermal cycler and incubated using the following conditions: initial denaturation (96°C - 1 min), followed by 25 cycles of denaturation (96°C - 10 sec), annealing (50°C - 5 sec) and extension (60°C - 75 sec).

During the denaturation stage, double-stranded templates are separated into single-stranded DNA (ssDNA) and during the annealing stage, primers hybridize to the ssDNA. At the extension

stage, the DNA polymerase incorporates dNTPs into growing chains of newly created DNA fragments that are complementary to the single-stranded template DNA. These extension products begin at the end of the primer and extend in the 3' direction and chain termination occurs during this extension step due to the incorporation of ddNTPs. Essentially, when a dNTP is incorporated, the new DNA fragment will continue to grow but when a ddNTP is incorporated, further chain elongation is blocked and this results in a population of truncated products of varying lengths. Finally, the double-stranded DNA fragments are denatured into ssDNA by adding 10μl of Hi-Di™ Formamide and incubating the mixture in thermal Cycler for 5 min.

The amplified and labelled DNA fragments are then electrophoresed to enable the separation of the labelled fragments and their visualization. This involves the loading of these fluorescently-labelled DNA fragments onto the ABI 3730xl capillary sequencer and during electrophoresis, the DNA fragments migrate through a polyacrylamide gel and are separated based on their size. Towards the end of the gel, the fluorescent dyes attached to the DNA fragments are excited by a laser and they emit light at a wavelength specific for each dye. This emitted light is separated according to wavelength by a spectrograph so that all four fluorescent emissions can be detected by one laser pass. The data collection software collects these light intensities at different wavelength bands and stores them onto the sequencer's computer as digital signals for processing. The analysis software then interprets the fluorescent intensity at each data point and assigns its base call interpretations and the data is finally output as chromatograms which contain the full sequence of the starting DNA fragment. These chromatograms are then analyzed using software for genomic sequence analysis.

4.2.4 Sequence analysis

Sequence assembly was performed in CLC Main Workbench v7.9.1 (Qiagen, UK) and SNPs were identified and called based on the respective 3D7 reference sequences. Nucleotide positions which displayed a peak within a peak in the sequence chromatograms were noted as ''mixed''. Consensus sequences were extracted from the sequence assemblies using CLC Genomics Workbench v.9.5.3 and used to construct multiple sequence alignments in Clustal

Omega v1.2.1 (Sievers and Higgins 2014). SNP frequencies were calculated per gene per time-point and singletons were confirmed by an additional round of PCR and sequencing.

4.2.5 Statistical analysis

Nucleotide sequences were translated into amino acid sequences in Aliview v1.26 (Larsson 2014). Haplotypes were then generated based on the amino acid residues from all the polymorphic codons that cut across all sequences and time points after excluding sequences with mixed bases. The difference in the prevalence of alleles and haplotypes in pre-ACT and post-ACT periods was evaluated using the Chi-squared test. For this analysis, 2005/06 was used as the point to divide the data into the pre- and post-ACT periods. Since complete ACT coverage is not expected following its formal introduction in 2004, data for 2005/06 was excluded from this part of the analysis. Chi-squared test for haplotypes was performed for only the sequences that had data across all loci at any time point. Additionally, the Chi-squared test for haplotypes was conducted only for the two dominant haplotypes. Linkage disequilibrium (LD) analysis was conducted in DnaSP v6.12.01 (Rozas *et al.* 2017) and reported based on the *D'* parameter that ranges between -1 and +1. Positive *D'* values imply that two alleles occur together on the same haplotype more often than expected, while negative values imply that alleles occur together on the same haplotype less often than expected (Lewontin 1964). All plots were generated using ggplot2 v3.1.1 (Wickham 2016) and ggpubr v0.2 packages (Kassambara 2020) in R v3.6.0 (R Core Team, 2014). Ka/Ks ratio was calculated for *k13* by dividing the number of nonsynonymous substitutions per non-synonymous site (Ka) by the number of synonymous substitutions per synonymous site over time.

4.3 Results

4.3.1 Genetic markers associated with artemisinin resistance

The *k13* codon K189T, was the only polymorphism maintained at frequencies (>10%, **Table 4.2**) and the dominant allele K189 was maintained at frequencies above 75% throughout the sampling period while the minor allele 189T had frequencies below 25% (**Figure 4.2**).

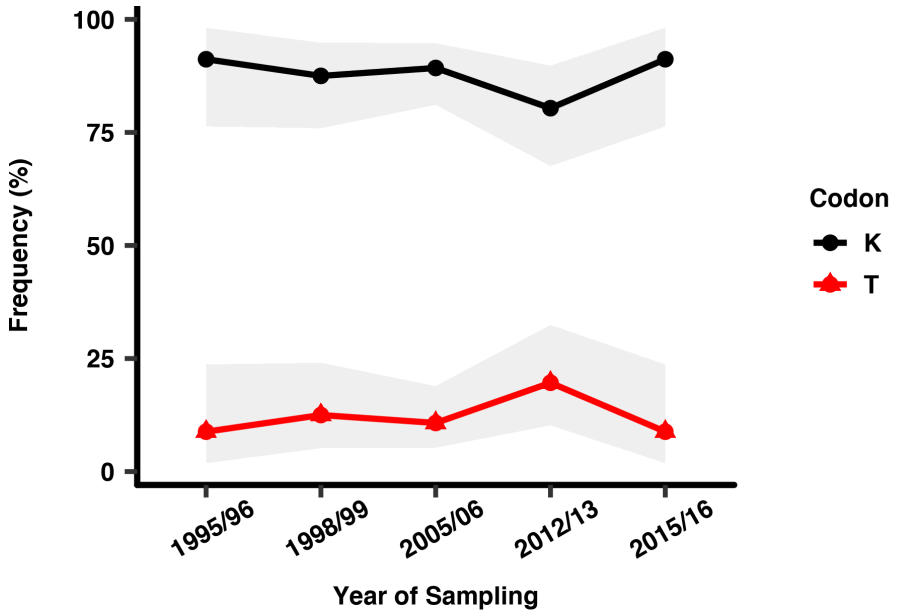


Figure 4.2. *k13* codon K189T frequencies over time. The two alleles appear to have stable frequencies from 1995/96 to 2005/06 but the frequency of K189 declines slightly to 2012/13 while the frequency of 189T increases slightly during the same period. After 2012/13, the frequencies of both alleles return to the frequencies observed in 1995/96. In grey are the 95% confidence intervals.

Other than K189T, the only other polymorphism observed across all time points is the asparagine (Asn) repeat at codon 137. This repeat region included insertions of between one and four asparagine residues, though at low frequencies <3% compared to SE Asia (>60%) (Wang *et al.* 2015b; Putaporntip *et al.* 2016). The rest of the observed *k13* alleles were rare including codon A578S, with frequencies barely reaching 2%. Many of the polymorphic codons occurred in the N-terminal region compared to the C-terminal region and from 1995/96 to 2015/16, the Ka/Ks ratio for the whole k13 gene ranged from 2.25 to 5. Conversely, the Ka/Ks ratio for the N-terminal region ranged from 2 to 9, while for the C-terminal region the Ka/Ks ratio was 1 throughout the same time period except for 2012/13 when there were no polymorphisms in the C-terminal region. The observations in the N-terminal region are comparable to those from other African studies, whereas fewer mutations were identified in SE Asian parasites (MalariaGEN Plasmodium falciparum Community Project 2016). Comparisons with the MalariaGEN dataset (MalariaGEN Plasmodium falciparum Community Project 2016) and other African studies (Torrentino-

Madamet *et al.* 2014; Boussaroque *et al.* 2016), revealed 9 loci that were unique to the Kilifi population and primarily occurred at only one time point over the 24 year study period. Codon K189T had similar frequencies to parasites examined in East Africa 13% and West Africa 50%, however the frequencies were much lower in SE Asia (<0.005%) (MalariaGEN Plasmodium falciparum Community Project 2016). Additionally, none of the *k13* mutations associated with resistance in SE Asia were identified and no SNPs were found to be in LD (**Table 4.2**).

Table 4.2. *k13* SNP frequencies

Codon	Nucleotide	Codon [Nucleotide]		Mutant Frequency % [n]				
		Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16
38	112	S [A]	C [T]	0 [132]	0 [117]	0 [133]	0 [114]	0.7 [135]
96	287	P [C]	Q [A]	0 [132]	0.7 [126]	0 [133]	0 [114]	0 [137]
108	322	K [A]	E [G]	0 [95]	0.7 [126]	0.7 [133]	0 [114]	0 [137]
119[a]	355	L [T]	L [C]	0.7 [132]	0 [126]	0.7 [133]	0 [114]	0 [136]
126[a]	377	T [C]	N [A]	0.7 [132]	0.7 [126]	0 [133]	0 [114]	0 [136]
134	401	I [T]	S [G]	0 [126]	0.7 [126]	0 [133]	0 [114]	0 [135]
136	406	H [C]	N [A]	0.7 [126]	0 [126]	0.7 [133]	1 [115]	0 [135]
137[a,b]	409	Nx6 [6xAAT]	Nx7 [7xAAT]	2 [124]	1 [126]	3 [133]	3 [115]	0.7 [134]
			Nx8 [8xAAT]	3 [124]	1 [126]	0 [133]	1 [115]	1 [134]
			Nx9 [9xTAA]	0 [124]	0 [126]	0 [133]	1 [115]	0 [134]
			Nx10 [10xTAA]	0 [124]	0 [126]	0 [133]	1 [115]	0 [134]
148	443	I [T]	T [C]	0 [122]	0 [104]	0.7 [133]	0 [114]	0 [133]
149[a]	445	T [A]	S [T]	0 [122]	0.7 [127]	0 [133]	0 [114]	1 [133]
157[a]	469	M [A]	V [G]	0 [122]	0.7 [127]	0 [133]	0 [114]	0 [132]
174	520	A [G]	S [T]	0 [95]	0 [125]	0 [133]	0.8 [115]	0 [83]
178[a]	532	I [A]	L [T]	1 [86]	0 [126]	0 [128]	0 [108]	0 [68]
182[a]	544	S [T]	T [A]	0 [95]	3 [125]	1 [132]	0 [117]	0 [87]
189[a]	566	K [A]	T [C]	8 [82]	15 [126]	10.6 [132]	15 [115]	13 [71]
	567	K [A]	T [T]	0 [79]	0 [124]	0.7 [132]	0.8 [115]	0 [71]
192[b]	574	T [A]	A [G]	0 [72]	0 [125]	0.7 [131]	0 [104]	0 [74]
258[a]	772	L [T]	M [A]	1 [91]	1 [100]	0 [105]	0 [79]	0.8 [116]
271[a]	813	Q [G]	H [T]	0 [93]	0 [69]	0.9 [107]	0 [81]	0 [121]
354	1060	I [A]	V [G]	0.7 [132]	0 [136]	0 [126]	0 [103]	0 [138]
417[a]	1251	P [C]	P [T]	0.7 [135]	0 [140]	0.7 [126]	0 [105]	0 [129]
469[a]	1407	C [C]	C [T]	0 [138]	0 [139]	2 [126]	0 [102]	0 [139]
487	1461	V [A]	V [T]	0 [139]	0.7 [127]	0 [130]	0 [105]	0 [142]
578[a,b]	1732	A [G]	S [T]	1 [137]	0.7 [127]	0.7 [127]	0 [117]	0.7 [142]
	1733	A [C]	S [T]	0.7 [137]	0 [131]	0 [127]	0 [117]	0 [142]
589[a]	1767	V [C]	V [T]	0 [136]	0 [123]	0 [127]	0 [117]	0.7 [139]

The number of samples successfully genotyped per timepoint include: 148 in 1995/96, 146 in 1998/99, 146 in 2005/06, 132 in 2012/13 and 148 in 2015/16. No sequences with mixed bases were identified. Frequency is presented as the percentage of sequences that carried a mutation out of the total number of sequences that had data for that locus [n].

Polymorphisms in codons 30 to 417 fall in the N-terminal region while those from 469 to 589 fall in the C-terminal region. In grey are zero frequencies and marked with [a] and [b] are N-terminal SNPs that have been identified in parasites from Africa and SE Asia, respectively. The rest of the SNPs appear to be unique to the Kilifi parasite population.

Of the 13 haplotypes, the 3D7 haplotype (PK[N6]MAISKLQ) was dominant over the entire sampling period with frequencies >70%. The second dominant haplotype (PK[N6]MAISTLQ), showed stable frequencies from 1995/96 to 2005/06 (8.8%-9.8%), however it roughly doubled to 19.6% in 2012/13 and then dropped back to 8.8% in 2015/16 (**Table 4.3**).

Table 4.3. *k13* haplotype frequencies

Haplotype	Frequency % [n]				
	1995/96	1998/99	2005/06	2012/13	2015/16
PK[N6]MAISKLQ*	76.4 [34]	75 [56]	82.9 [94]	71.4 [56]	88.2 [34]
PK[N6]MAISTLQ	8.8 [34]	12.5 [56]	9.5 [94]	19.6 [56]	8.8 [34]
PK[N7]MAISKLQ	5.8 [34]	1.7 [56]	2.1 [94]	3.5 [56]	2.9 [34]
PK[N6]MAISKMQ	2.9 [34]	1.7 [56]	0 [94]	0 [56]	0 [34]
PK[N6]MALSKLQ	2.9 [34]	0 [56]	0 [94]	0 [56]	0 [34]
PK[N8]MAISKLQ	2.9 [34]	0 [56]	0 [94]	3.5 [56]	0 [34]
PK[N6]MAITKLQ	0 [34]	5.3 [56]	2.1 [94]	0 [56]	0 [34]
PK[N6]VAISKLQ	0 [34]	1.7 [56]	0 [94]	0 [56]	0 [34]
QK[N6]MAISKLQ	0 [34]	1.7 [56]	0 [94]	0 [56]	0 [34]
PE[N6]MAISKLQ	0 [34]	0 [56]	1.0 [94]	0 [56]	0 [34]
PK[N6]MAISNLQ	0 [34]	0 [56]	1.0 [94]	0 [56]	0 [34]
PK[N6]MAISTLH	0 [34]	0 [56]	1.0 [94]	0 [56]	0 [34]
PK[N6]MSISKLQ	0 [34]	0 [56]	0 [94]	1.7 [56]	0 [34]

The following were used to define haplotypes: 10 polymorphic sites, including codons 96, 108, 137, 157, 174, 178, 182, 189, 258 and 271. The haplotypes marked with an * indicate the 3D7 haplotype. Frequency is presented as the percentage of sequences that contributed to the haplotype out of the total number of sequences [n]. In grey are zero SNP frequencies.

4.3.2 Other putative genetic markers associated with artemisinin resistance

There were no mutations identified in *arps10*, *crt* and *fd*. However, a high frequency (>10%) SNP (I492V) was identified in *mdr2* in the 1995/96 and 2015/16 time points and hence the three middle time points were included. Consequently, two additional polymorphisms were identified (I495V – 0.9% in 1998/99 and V506I – 0.7% in 2005/06) with the I492V polymorphism maintaining high frequencies (between 13% to 30%, **Table 4.4**).

Table 4.4. *mdr2* SNP frequencies

Codon	Nucleotide	Codon [Nucleotide]		Mutant Frequency % [n]				
		Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16
492	1474	I [A]	V [G]	30.5 [131]	21 [108]	13 [132]	14 [116]	23 [130]
495	1483	I [A]	V [G]	0 [131]	0.9 [108]	0 [139]	0 [116]	0 [130]
506	1516	V [G]	I [A]	0 [129]	0 [107]	0.71 [136]	0 [115]	0 [130]

Supplementary Table 2. *Mdr2* SNP frequencies. The number of samples successfully genotyped per timepoint include: 131 in 1995/96, 117 in 1998/99, 134 in 2005/06, 145 in 2012/13 and 130 in 2015/16. No sequences with mixed bases were identified. Frequency is presented as the percentage of sequences that carried a mutation out of the total number of sequences that had data for that locus [n]. In grey are zero SNP frequencies.

In *ap2-mu*, the following polymorphisms were observed across all time points, I100I, S160N, E163E, repeat-[7N]227[6N/8N/9N/10N], repeat-[5N]320[6N/7N], S476S and V478V. In addition, 2 other repeat blocks were identified, [K]234[2K], and [4N]327[5N/6N]. All the repeat blocks had stable frequencies over time except [K]234[2K], that was at a low frequency of (~2%), however it was not detected in 2005/06, was later detected in 2012/13 at 10.7% and later dropped to 1.8% in 2015/16. No significant temporal trends were observed and only two polymorphisms achieved frequencies >10% across time, I100I and [7N]227[6N/8N/9N/10N]. Over the same period, the synonymous mutation E163E showed similar increases and decreases in frequency from 8.51% in 2006/06 to 16.36% in 2012/13 and down to 4.55% in 2015/16) (**Table 4.5**).

Table 4.5. *ap2-mu* SNP frequencies

Codon	Nucleotide	Codon [Nucleotide]		Mutant Frequency % [n]				
		Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16
44	132	E [A]	E [G]	0 [93]	0 [110]	2 [48]	0 [43]	0 [77]
92	275	F [T]	S [C]	0 [87]	0 [110]	2 [48]	0 [48]	0 [81]
100	300	I [A]	I [C]	29 [86]	30 [106]	37 [37]	36 [46]	31 [86]
127	381	V [G]	V [A]	0 [74]	0 [101]	0 [21]	2 [49]	0 [92]
144	430	T [A]	A [G]	0 [75]	0 [102]	3 [27]	0 [51]	0 [90]
146	437	R [G]	K [A]	0 [74]	1 [102]	3 [27]	0 [51]	0 [91]
149	447	Q [G]	Q [A]	1 [76]	1 [101]	3 [27]	0 [52]	2 [92]
160	479	S [G]	N [A]	10 [75]	11 [102]	9 [44]	24 [44]	16 [109]
162	486	I [T]	I [C]	0 [77]	3 [103]	0 [47]	0 [55]	1 [110]
163	489	E [A]	E [G]	7 [77]	12 [102]	8 [47]	16 [55]	4 [110]
188	564	R [A]	R [G]	1 [85]	0 [104]	0 [51]	0 [57]	0 [117]
199	596	K [A]	T [C]	11 [86]	6 [102]	0 [59]	16 [60]	11 [114]
	597	K [A]	K [G]	0 [86]	0 [102]	3 [59]	0 [60]	0.8 [114]
200	598	N [A]	Y [T]	1 [87]	0 [102]	1 [60]	0 [60]	0 [114]

207	620	T [C]	R [G]	1 [89]	0.9 [102]	0 [59]	0 [60]	0 [115]
225	675	I [A]	I [T]	0 [90]	0 [101]	0 [57]	1 [63]	0 [114]
227	679	Nx7 [7xAAT)	Nx6 [6xAAT]	0 [81]	0.9 [81]	0 [57]	0 [65]	0 [116]
			Nx8 [8xAAT]	10 [81]	13.7 [81]	17 [57]	12 [65]	14 [116]
			Nx9 [9xAAT]	6 [81]	5.8 [81]	7 [57]	1 [65]	12 [116]
			Nx10 [10xAAT]	0 [81]	0.9 [81]	0 [57]	0 [65]	0.8 [116]
234	700	Kx1 [AAG]	Kx2 [2xAAG]	2 [91]	2 [101]	0 [57]	10 [65]	1 [111]
235	704	T [C]	I [T]	1 [89]	0 [100]	0 [52]	0 [65]	0 [112]
236	706	A [G]	T [A]	1 [88]	0 [100]	1 [51]	1 [66]	0 [112]
254	760	Y [T]	H [C]	1 [85]	0 [102]	0 [51]	0 [65]	0 [112]
274	821	S [G]	I [T]	0 [83]	0 [104]	2 [47]	0 [66]	0 [109]
289	867	K [A]	K [G]	1 [82]	0 [99]	0 [47]	0 [66]	0 [109]
315	943	G [G]	S [A]	0 [85]	0 [96]	1 [52]	0 [70]	0 [111]
	944	G [G]	D [A]	1 [85]	1 [96]	0 [52]	0 [70]	0 [111]
320	958	Nx5 [AAC(4xAAT)]	Nx6 [AAC(5xAAT)]	6 [81]	2 [76]	10 [52]	7 [69]	3 [105]
			Nx7 [AAC(6xAAT)]	0 [81]	2 [76]	0 [52]	0 [69]	0 [106]
327	979	Nx4 [(3xAAT)AAC]	Nx5 [(4xAAT)AAC]	2 [75]	1 [84]	2 [50]	0 [66]	2 [107]
			Nx6 [(5xAAT)AAC]	0 [75]	0 [84]	0 [50]	0 [66]	0.9 [107]
337	1009	A [G]	T [A]	0 [78]	1 [81]	0 [46]	1 [71]	0 [113]
341	1023	S [T]	S [C]	0 [77]	1 [81]	0 [45]	2 [69]	0 [113]
437	1311	F [C]	L [A]	9 [72]	0 [89]	4 [65]	4 [97]	1 [118]
445	1334	N [A]	S [G]	0 [75]	0 [91]	1 [66]	0 [97]	0 [118]
476	1428	S [T]	S [G]	1 [79]	6 [99]	2 [79]	0.9 [101]	1.6 [121]
478	1434	V [A]	V [T]	2 [78]	4 [103]	2 [79]	1.9 [101]	2.4 [121]
491	1473	V [T]	V [T]	0 [80]	0 [96]	0 [76]	0 [100]	0.8 [122]
498	1493	I [T]	T [C]	0 [80]	2 [95]	0 [74]	0 [98]	0 [121]
517	1551	F [C]	F [T]	0 [84]	1 [95]	0 [73]	0 [94]	0 [119]
548	1644	D [C]	D [T]	0 [84]	0 [86]	0 [66]	0 [89]	0.9 [111]
586	1758	T [C]	T [A]	0 [31]	0 [64]	0 [58]	0 [79]	0.9 [104]
603	1809	Y [T]	Y [C]	0 [4]	0 [47]	1 [54]	0 [72]	0 [100]

The number of samples successfully genotyped per timepoint include: 130 in 1995/96, 134 in 1998/99, 123 in 2005/06, 110 in 2012/13 and 135 in 2015/16. There were a total of 13, 16, 3,3 and 13 sequences with mixed bases in 1995/96, 1998/99, 2005/06, 2012/13 and 2015/16, respectively. Frequency is presented as the percentage of sequences that carried a mutation out of the total number of sequences that had data for that locus [n]. In grey are zero SNP frequencies.

No SNPs were found to be in LD and the prevalence of the S160N mutation was similar from 1995/96 to 2015/16, except for a two-fold increase from 9% in 1998/99 to 24% in 2012/13 that later decreased to 16% in 2015/16 (**Figure 4.3**).

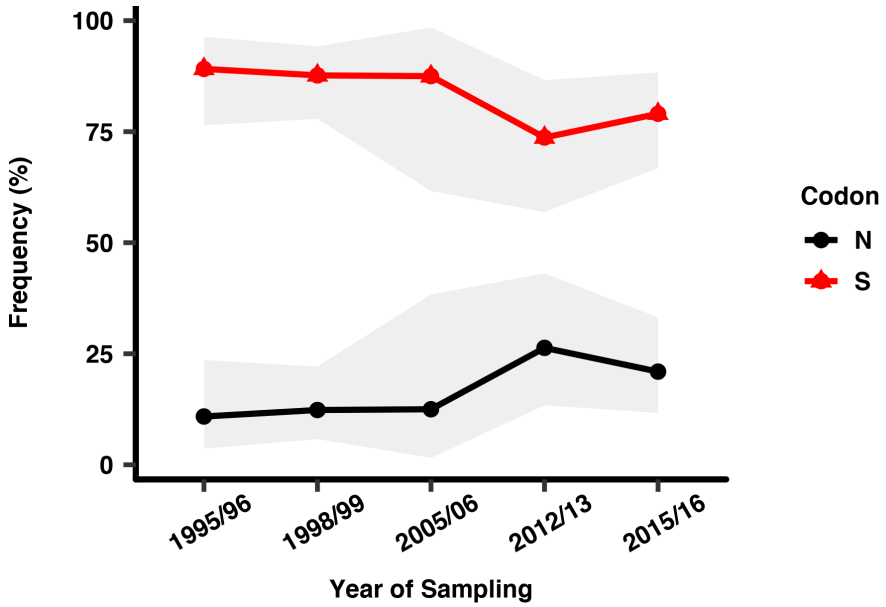


Figure 4.3. *ap2-mu* codon S160N frequencies over time. The two alleles appear to have stable frequencies from 1995/96 to 2005/06 but the frequency of S160 declines to 2012/13 while the frequency of 160N increases during the same period. After 2012/13, the frequencies of S160 increases and that of 160N decreases. In grey are the 95% confidence intervals.

A total of 38 *ap2-mu* haplotypes were assembled with the 3D7 haplotype TRSKT[N7][Kx1]AG[N5][N4]AFI dominating across time (>39%, **Table 4.6**).

Table 4.6. *ap2-mu* haplotype frequencies

Haplotype	Frequency % [n]				
	1995/96	1998/99	2005/06	2012/13	2015/16
TRSKT[N7][Kx1]AG[N5][N4]AFI*	52.1 [46]	56.1 [73]	43.7 [16]	39.4 [38]	40.3 [62]
TRNKT[N7][Kx1]AG[N5][N4]AFI	6.5 [46]	4.1 [73]	0 [16]	13.1 [38]	11.2 [62]
TRSKT[N7][Kx1]AG[N5][N4]ALI	6.5 [46]	0 [73]	0 [16]	5.2 [38]	3.2 [62]
TRSKT[N9][Kx1]AG[N5][N4]AFI	6.5 [46]	5.4 [73]	6.2 [16]	0 [38]	9.6 [62]
TRSKT[N7][Kx1]AG[N5][N5]AFI	4.3 [46]	0 [73]	0 [16]	0 [38]	3.2 [62]
TRSKT[N7][Kx1]AG[N6][N4]AFI	4.3 [46]	2.7 [73]	6.2 [16]	0 [38]	0 [62]
TRSKT[N8][Kx1]AG[N5][N4]AFI	4.3 [46]	8.2 [73]	18.7 [16]	7.8 [38]	11.2 [62]
TRSTT[N7][Kx1]AG[N5][N4]AFI	4.3 [46]	2.7 [73]	0 [16]	5.2 [38]	1.6 [62]
TRNKT[N8][Kx1]AG[N5][N4]ALI	2.1 [46]	0 [73]	0 [16]	0 [38]	0 [62]
TRNKT[N8][Kx1]TG[N5][N4]AFI	2.1 [46]	0 [73]	0 [16]	0 [38]	0 [62]
TRSKR[N7][Kx1]AG[N6][N4]AFI	2.1 [46]	0 [73]	0 [16]	0 [38]	0 [62]
TRSTT[N7][Kx2]AG[N5][N4]AFI	2.1 [46]	0 [73]	0 [16]	0 [38]	0 [62]
TRSTT[N9][Kx1]AG[N5][N4]AFI	2.1 [46]	0 [73]	0 [16]	0 [38]	1.6 [62]
TRNKT[N7][Kx1]AG[N6][N4]AFI	0 [46]	2.7 [73]	0 [16]	0 [38]	0 [62]
TRSTT[N8][Kx1]AG[N5][N4]AFI	0 [46]	2.7 [73]	0 [16]	0 [38]	3.2 [62]
TKSKT[N9][Kx1]AG[N5][N4]AFI	0 [46]	1.3 [73]	6.2 [16]	0 [38]	0 [62]

TRNKT[N7][Kx1]AG[N5][N5]AFI	0 [46]	1.3 [73]	0 [16]	0 [38]	0 [62]
TRNKT[N8][Kx1]AG[N5][N4]AFI	0 [46]	1.3 [73]	6.2 [16]	0 [38]	1.6 [62]
TRNKT[N8][Kx1]AG[N5][N4]AFT	0 [46]	1.3 [73]	0 [16]	0 [38]	0 [62]
TRNTT[N7][Kx1]AG[N5][N4]AFI	0 [46]	1.3 [73]	0 [16]	2.6 [38]	4.8 [62]
TRSKT[N10][Kx1]AG[N5][N4]AFI	0 [46]	1.3 [73]	0 [16]	0 [38]	0 [62]
TRSKT[N6][Kx2]AG[N5][N4]TFI	0 [46]	0 [73]	0 [16]	0 [38]	0 [62]
TRSKT[N7][Kx1]AG[N7][N4]AFI	0 [46]	1.3 [73]	0 [16]	0 [38]	0 [62]
TRSKT[N7][Kx2]AG[N6][N4]AFI	0 [46]	1.3 [73]	0 [16]	2.6 [38]	0 [62]
TRSKT[N9][Kx1]AD[N5][N4]AFI	0 [46]	1.3 [73]	0 [16]	0 [38]	0 [62]
TRSTT[N7][Kx1]AG[N7][N4]AFI	0 [46]	1.3 [73]	0 [16]	0 [38]	0 [62]
ARSKT[N8][Kx1]AG[N5][N4]AFI	0 [46]	0 [73]	6.2 [16]	0 [38]	0 [62]
TRNKT[N9][Kx1]AG[N5][N4]AFI	0 [46]	0 [73]	6.2 [16]	0 [38]	0 [62]
TRNKT[N7][Kx2]AG[N5][N4]AFI	0 [46]	0 [73]	0 [16]	10.5 [38]	0 [62]
TRSKT[N7][Kx1]AG[N5][N4]TFI	0 [46]	0 [73]	0 [16]	2.6 [38]	0 [62]
TRSKT[N7][Kx2]AG[N5][N4]AFI	0 [46]	0 [73]	0 [16]	2.6 [38]	0 [62]
TRSKT[N8][Kx1]AG[N6][N4]AFI	0 [46]	0 [73]	0 [16]	2.6 [38]	0 [62]
TRSTT[N8][Kx1]AG[N6][N4]AFI	0 [46]	0 [73]	0 [16]	2.6 [38]	0 [62]
TRSTT[N9][Kx1]AG[N6][N4]AFI	0 [46]	0 [73]	0 [16]	2.6 [38]	0 [62]
TRNTT[N9][Kx1]AG[N5][N4]AFI	0 [46]	0 [73]	0 [16]	0 [38]	3.2 [62]
TRSKT[N10][Kx1]AG[N5][N6]AFI	0 [46]	0 [73]	0 [16]	0 [38]	1.6 [62]
TRSKT[N9][Kx1]AG[N6][N4]AFI	0 [46]	0 [73]	0 [16]	0 [38]	1.6 [62]
TRSTT[N7][Kx1]AG[N6][N4]AFI	0 [46]	0 [73]	0 [16]	0 [38]	1.6 [62]

The following were used to define haplotypes: 14 polymorphic sites, including codons 144, 146, 160, 199, 207, 227, 234, 236, 315, 320, 327, 337, 437, 498. The haplotypes marked with an * indicate the 3D7 haplotype. Frequency is presented as the percentage of sequences that contributed to the haplotype out of the total number of sequences [n]. In grey are zero SNP frequencies.

falcipain-2a was found to be the most polymorphic gene, however, the S69Stop polymorphism was not identified. The following polymorphic codons were observed across all time points N4H, A8I, A8I, P9P, H10N, E11E, Q15H, V21V, K31K, S43S, V47V, V51I, S59F, N173K, S228T, K255R, N257E, F260F, Y265Y, D266D, H270H, M335I, G339G, T343P, D345G, V350V, N358N, S373S, Q414E and L474L and had frequencies >5% (**Table 4.7**). Additionally, many SNPs under 1250bp were found to be in high LD (**Figure 4.4**).

Table 4.7. *falcipain-2a* SNP frequencies

Codon	Nucleotide	Codon [Nucleotide]		Mutant Frequency % [n]				
		Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16
4	10	N [A]	H [C]	9 [91]	9 [113]	19 [73]	9 [98]	16 [109]
8	22	A [G]	I [A]	8 [93]	7 [113]	16 [74]	8 [100]	14 [110]
	23	A [C]	I [T]	8 [93]	11 [113]	17 [74]	8 [100]	15 [110]
9	27	P [C]	P [G]	8 [94]	7 [113]	17 [74]	6 [100]	14 [110]
10	28	H [C]	N [A]	8 [96]	8 [112]	17 [74]	7 [99]	14 [110]
11	33	E [A]	E [G]	8 [96]	9 [112]	18 [74]	7 [99]	15 [111]

15	45	Q [A]	H [T]	10 [98]	13 [113]	14 [78]	13 [99]	13 [113]
21	63	V [T]	V [C]	11 [98]	15 [113]	18 [79]	13 [101]	15 [112]
31	93	K [G]	K [A]	2 [99]	6 [116]	3 [82]	1 [101]	0.8 [116]
43	129	S [T]	S [A]	10 [100]	8 [116]	5 [81]	4 [102]	4 [116]
47	141	V [T]	V [A]	12 [101]	11 [115]	11 [80]	8 [103]	8 [115]
48	142	V [G]	I [A]	0 [142]	0 [115]	0 [81]	0 [103]	0.8 [115]
51	151	V [G]	I [A]	13 [104]	13 [115]	13 [81]	9 [104]	10 [118]
59	176	S [C]	F [T]	12 [105]	8 [116]	7 [80]	1 [104]	5 [118]
60	179	R [G]	K [A]	0 [105]	0 [117]	0 [79]	0.9 [102]	0.8 [119]
82	244	L [T]	L [C]	0.9 [106]	0.8 [117]	0 [81]	0.9 [103]	0 [118]
85	255	S [C]	S [T]	0 [106]	0 [119]	0 [81]	0.9 [102]	0 [118]
86	256	P [C]	T [A]	0 [106]	0 [119]	0 [81]	0 [103]	0 [118]
87	260	N [A]	N [C]	0 [106]	0.8 [119]	0 [79]	0 [102]	0 [118]
114	340	N [A]	D [G]	0 [103]	0 [120]	1 [83]	0 [101]	0 [118]
	342	N [C]	N [T]	1 [103]	0 [120]	0 [83]	0 [101]	0 [118]
115	344	E [A]	V [T]	2 [103]	3 [120]	1 [86]	0 [101]	0.8 [118]
123	369	D [C]	D [T]	0.9 [105]	0 [120]	0 [86]	0 [102]	0 [119]
126	377	G [G]	D [A]	0 [105]	3 [122]	0 [86]	0 [103]	0 [120]
127	381	L [T]	L [A]	0.9 [106]	0 [122]	0 [86]	0.9 [103]	0 [120]
134	401	T [C]	K [A]	0.9 [105]	0 [123]	0 [89]	0.9 [106]	0 [120]
140	418	I [A]	L [C]	0.9 [106]	0 [124]	0.9 [105]	0 [108]	0 [120]
143	427	K [A]	E [G]	0 [107]	0 [125]	0.9 [105]	1 [108]	0 [120]
144	430	D [G]	N [A]	0 [106]	0 [125]	0.9 [105]	1 [108]	0 [120]
150	449	F [T]	Y [A]	0 [106]	0.8 [125]	0 [105]	0 [111]	0 [120]
167	500	M [T]	T [C]	0.9 [107]	0.8 [122]	0 [98]	0 [112]	0 [120]
169	507	I [T]	I [A]	0.9 [106]	0.8 [122]	0 [98]	0 [112]	0 [120]
173	519	N [T]	K [A]	1 [105]	2 [122]	1 [98]	0.8 [112]	4 [120]
179	537	P [A]	P [G]	0 [106]	2 [123]	2 [98]	6 [112]	2 [118]
183	549	K [G]	K [A]	0 [106]	0 [122]	1 [98]	0 [114]	0 [119]
204	612	N [T]	K [G]	0 [104]	0.8 [122]	0 [98]	0 [113]	0 [122]
210	628	E [G]	Q [C]	0 [104]	0 [122]	0 [98]	0.8 [114]	4 [121]
212	636	N [C]	N [T]	0 [104]	0.8 [122]	0 [97]	0 [114]	0 [121]
215	645	A [C]	A [T]	0 [104]	0.8 [124]	0 [97]	0 [113]	0 [121]
218	654	T [T]	T [A]	0 [105]	0 [124]	0 [97]	0.8 [113]	0.8 [121]
220	660	H [T]	H [C]	0 [105]	0.8 [124]	0 [97]	0 [113]	0 [121]
224	671	N [A]	S [G]	0 [105]	0.8 [124]	0 [97]	1 [113]	0 [121]
	672	N [C]	S [T]	0 [105]	0.8 [124]	0 [97]	2 [113]	0 [121]
227	681	L [T]	L [A]	0.9 [105]	0 [124]	0 [97]	3 [113]	0 [120]
228	683	S [G]	T [C]	10 [105]	13 [124]	10 [97]	7 [113]	5 [121]
232	696	S [A]	S [G]	0 [105]	1 [124]	0 [96]	2 [113]	0 [120]
245	735	M [G]	I [A]	0 [107]	3 [126]	0 [95]	2 [113]	0 [120]
248	744	E [A]	D [C]	0 [106]	1 [126]	0 [95]	2 [113]	0 [119]
249	746	E [A]	A [C]	0 [106]	1 [126]	0 [95]	2 [113]	0 [119]
	747	E [A]	A [C]	0 [106]	1 [126]	0 [96]	2 [113]	0 [119]
250	750	V [T]	V [A]	0 [106]	1 [125]	0 [96]	2 [112]	0 [119]
255	763	K [A]	G [G]	1 [106]	0.8 [125]	0 [97]	0.9 [111]	0.8 [118]
	764	K [A]	R [G]	92 [106]	97 [125]	97 [108]	98 [111]	1 [118]
257	769	N [A]	E [G]	92 [106]	96 [124]	97 [95]	96 [111]	97 [118]

	771	N [T]	E [A]	92 [106]	96 [124]	97 [95]	96 [111]	97 [118]
260	780	F [T]	F [C]	92 [106]	96 [123]	97 [95]	96 [111]	96 [118]
265	795	Y [T]	Y [C]	92 [106]	96 [121]	97 [92]	96 [110]	96 [119]
266	798	D [T]	D [C]	92 [106]	96 [121]	97 [92]	95 [110]	95 [120]
270	810	H [T]	H [C]	92 [100]	96 [123]	97 [93]	98 [110]	98 [120]
299	896	A [C]	V [T]	2 [112]	3 [122]	2 [93]	0 [106]	0 [120]
321	962	Y [A]	S [C]	0 [108]	0.8 [120]	1 [93]	0 [94]	0 [121]
335	1005	M [G]	I [T]	2 [108]	0.8 [119]	1 [92]	1 [97]	1 [121]
339	1017	G [G]	G [A]	88 [108]	88 [120]	92 [100]	88 [97]	89 [121]
343	1027	T [A]	P [C]	90 [108]	89 [119]	93 [100]	88 [98]	90 [121]
345	1034	D [A]	G [G]	90 [108]	89 [119]	94 [100]	92 [98]	90 [122]
350	1048	V [G]	M [A]	0 [109]	0.8 [119]	0 [99]	0 [99]	0 [122]
350	1050	V [A]	V [G]	90 [109]	89 [119]	6 [96]	93 [99]	9 [122]
358	1074	N [T]	N [C]	91 [112]	86 [121]	96 [96]	10 [99]	10 [123]
373	1119	S [C]	S [T]	8 [114]	11 [123]	3 [96]	6 [7]	9 [124]
376	1127	D [A]	V [T]	0 [114]	0.7 [126]	0 [102]	0 [103]	0 [124]
393	1177	V [G]	I [A]	0.8 [116]	0 [126]	0 [101]	0 [104]	2 [124]
400	1198	A [G]	P [C]	0.8 [116]	0 [127]	0 [101]	1 [104]	2 [124]
	1200	A [T]	P [G]	0.8 [116]	0 [127]	0 [101]	2 [104]	2 [124]
407	1221	F [C]	F [T]	0.8 [116]	0 [128]	0 [101]	2 [105]	2 [123]
414	1240	Q [C]	E [G]	48 [/116]	42 [128]	33 [34]	41 [105]	41 [123]
434	1302	T [C]	T [T]	0 [114]	0 [128]	1 [99]	0.8 [122]	0 [122]
474	1420	L [T]	L [C]	0.9 [103]	0.8 [125]	3 [99]	0.9 [110]	0 [110]

The number of samples successfully genotyped per timepoint include: 137 in 1995/96, 141 in 1998/99, 120 in 2005/06, 125 in 2012/13 and 126 in 2015/16. There were a total of 17, 7, 11,3 and 0 sequences with mixed bases in 1995/96, 1998/99, 2005/06, 2012/13 and 2015/16, respectively. Frequency is presented as the percentage of sequences that carried a mutation out of the total number of sequences that had data for that locus [n]. In grey are zero SNP frequencies.

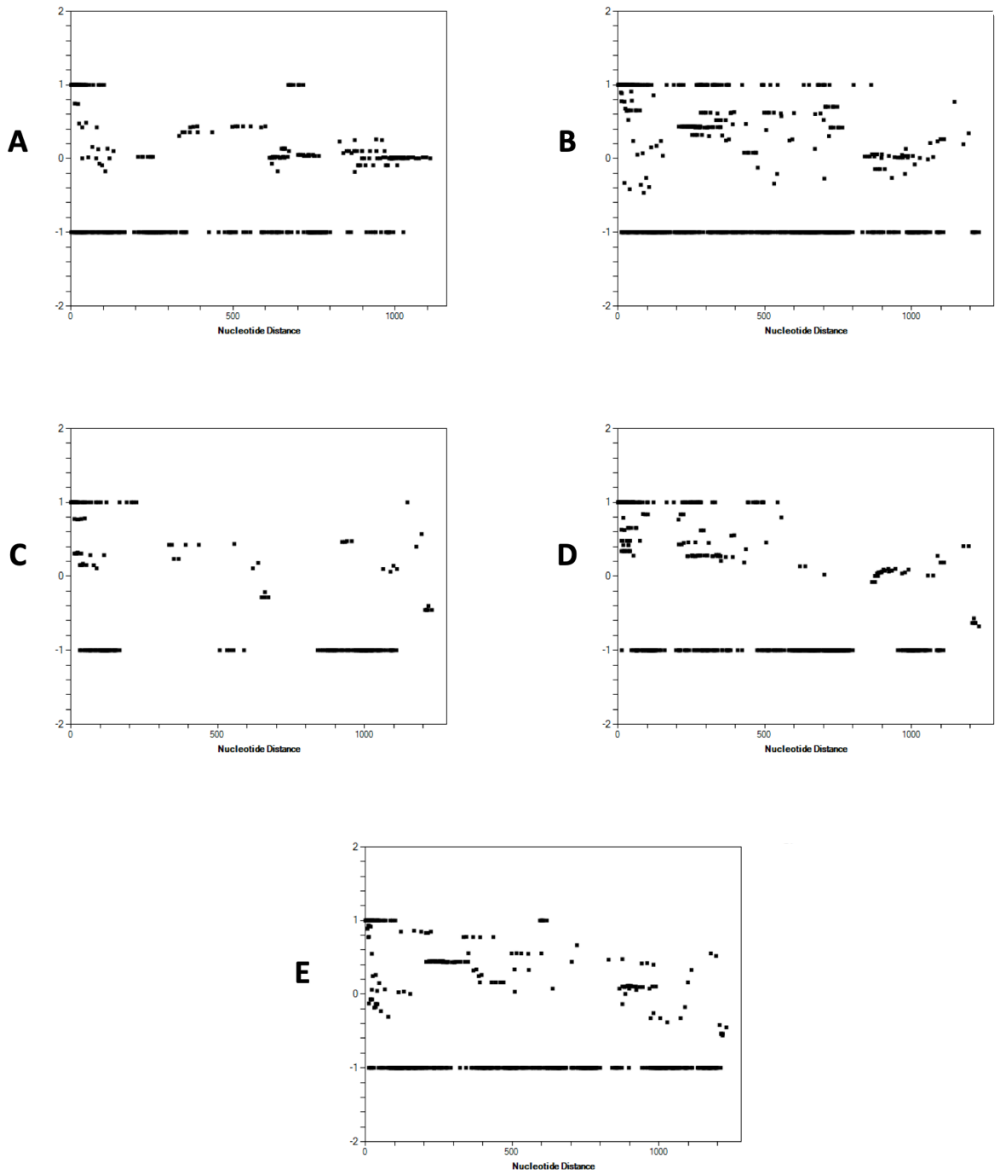


Figure 4.4. Linkage disequilibrium (LD) scatter plot for falcipain-2a. (A) 1995/96, (B) 1998/99, (C) 2005/06, (D) 2012/13, (D) 2015/16. LD or non-random association between alleles was estimated with the D' parameter (y-axis) and plotted against the nucleotide distance (x-axis). The black dots represent the D' value for each of the pairwise comparisons conducted at each time point. Over the entire sampling period, many SNPs under 1250 bp were in high LD (D' > 0).

A total of 71 *falcipain-2a* haplotypes were assembled and only two haplotypes exceeded 10% frequency across time (NAQSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ and NAQSRNEGTIKDFMNNENSMEEREAYMPGVDVAE). Many of the haplotypes were rare, appearing and disappearing at different time points, while the 3D7 haplotype did not occur

at any time point (**Table 4.8**). The only significant temporal trend observed was that of codon S59F ($\chi 2$ = 5.7, p-value = 0.02), with the 59F allele dropping from 13% in 1995/96 to 5% in 2015/16.

Table 4.8. *falcipain-2a* haplotype frequencies

Haplotype	Frequency % [n]				
	1995/96	1998/99	2005/06	2012/13	2015/16
NAQSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	22.8 [83]	31.7 [104]	37.3 [67]	45.7 [83]	33.6 [104]
NAQSRNEGTIKDFMNNENSMEEREAYMPGVDVAE	20.4 [83]	14.4 [104]	10.4 [67]	14.4 [83]	15.3 [104]
HIQSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	4.8 [83]	7.6 [104]	10.4 [67]	3.6 [83]	5.7 [104]
NAQFRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	4.8 [83]	1.9 [104]	4.4 [67]	1.2 [83]	4.8 [104]
NAQSRNEGTIKDFMNNENTMEEREAYMTDVDVAE	4.8 [83]	6.7 [104]	1.4 [67]	2.4 [83]	1.9 [104]
NAQSRNEGTIKDFMNNENSMEEKNAYMPGVDVAE	3.6 [83]	0 [104]	1.4 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMNNENSMEEKNAYMPGVDVAQ	3.6 [83]	1.9 [104]	0 [67]	0 [83]	0.9 [104]
NAQSRNEGTIKDFMNNENTMEEREAYMPGVDVAQ	3.6 [83]	1.9 [104]	2.9 [67]	1.2 [83]	0 [104]
HIQSRNEGTIKDFMNNENTMEEREAYMPGVDVAQ	2.4 [83]	0 [104]	1.4 [67]	0 [83]	0 [104]
NAQFRNEGTIKDFMNNENSMEEREAYMPGVDVAE	2.4 [83]	0.9 [104]	2.9 [67]	0 [83]	0 [104]
HAHSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	1.2 [83]	0 [104]	1.4 [67]	1.2 [83]	0 [104]
HIQSRNEGTIKDFMNNENSMEEREAYMPGVDVAE	1.2 [83]	0 [104]	0 [67]	0 [83]	1.9 [104]
HIQSRNEGTIKDFMNNENSMEEREAYMTDVDVAE	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAHFRNEGTIKDFMNNENSMEEREAYMPGVDVAE	1.2 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NAHFRNEGTIKDFMNNENSMEEREAYMTDVDVAE	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAHSRNEGKIKDFMNNENSMEEREAYMPGVDVAE	1.2 [83]	0 [104]	0 [67]	1.2 [83]	0 [104]
NAHSRNEGTIKDFMNNENSMEEKNAYMPGVDVAQ	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAHSRNEGTIKDFMNNENSMEEREAYIPGVDVAE	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAHSRNEGTIKDFMNNENSMEEREAYMPGVDVAE	1.2 [83]	4.8 [104]	2.9 [67]	3.6 [83]	7.6 [104]
NAHSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	1.2 [83]	0.9 [104]	0 [67]	1.2 [83]	0.9 [104]
NAHSRNEGTIKDFMNNENTMEEREAYMPGVDVAE	1.2 [83]	0 [104]	2.9 [67]	2.4 [83]	0 [104]
NAQFRNEGTIKDFMNNENSMEEGEAYMPGVDVAQ	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAQFRNEGTIKDFMNNENSMEEREAYIPGVDVAQ	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAQFRNEGTIKDFMNNENSMEEREAYMTDVDVAE	1.2 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMKNENSMEEREAYMPGVDVAQ	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMKNENSMEEREAYMTDVDVAE	1.2 [83]	1.9 [104]	1.4 [67]	1.2 [83]	1.9 [104]
NAQSRNEGTIKDFMNNENSMEEREAYIPGVDVAQ	1.2 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
NAQSRNEGTIKDFMNNENSMEEREAYMTDVDVAE	1.2 [83]	0 [104]	1.4 [67]	2.4 [83]	2.8 [104]
NAQSRNEGTIKDFMNNENSMEEREVYMPGVDVAE	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMNNENTMEEGEAYMPGVDVAE	1.2 [83]	0.9 [104]	0 [67]	1.2 [83]	0 [104]
NAQSRNEGTIKDFTNNENSMEEREAYMPGVDVAQ	1.2 [83]	0 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTLKDFMNNENSMEEREAYMPGVDVAQ	1.2 [83]	0 [104]	1.4 [67]	0 [83]	0 [104]
NAQSRNEDTIKDFMNNENSMEEREVYMPGVDVAE	0 [83]	3.8 [104]	0 [67]	0 [83]	0 [104]
NAHSRNVGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	2.8 [104]	0 [67]	0 [83]	0.9 [104]
NAHSRNEGTIKDFMNNENSIDAKNAYMTDVDVAE	0 [83]	1.9 [104]	0 [67]	0 [83]	0 [104]
HAQSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
HVQSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0.9 [104]	1.4 [67]	0 [83]	0 [104]
NAHSRNEGTIKDFMNNENSMEEREAYMPGMDVAE	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]

NAQFRNEGTIKDFMNNENSMEEREAYIPGVDVAE	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMKNENSMEEREAYMPGVDVAE	0 [83]	0.9 [104]	0 [67]	0 [83]	0.9 [104]
NAQSRNEGTIKDFMNKESTIEEREAYMPGVDVAQ	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMNNENSMEEKEAYMPGVDVAQ	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMNNENSMEEREASMPGVDVAE	0 [83]	0.9 [104]	1.4 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMNNENSMEEREAYMPGVVVAE	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMNNENTMEEREAYMPGVDVAE	0 [83]	0.9 [104]	1.4 [67]	1.2 [83]	0 [104]
NAQSRNEGTIKDFTNNENTMEEREAYMPGVDVAE	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NAQSRNEGTIKDYMNNENSMEEREAYMPGVDVAE	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NVHFRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
NVQSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0.9 [104]	0 [67]	0 [83]	0 [104]
HIHSRNEGTIKDFMNNENSMEEREAYMPGVDVAE	0 [83]	0 [104]	2.9 [67]	1.2 [83]	0.9 [104]
HIHSRNEGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0 [104]	1.4 [67]	2.4 [83]	0.9 [104]
NAHSRDEGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0 [104]	1.4 [67]	0 [83]	0 [104]
NAQSRNEGTIENFMNNENSMEEREAYMPGVDVAQ	0 [83]	0 [104]	1.4 [67]	1.2 [83]	0 [104]
NAQSRNEGTIKDFMNNENTMEEREAYMTGVDVAE	0 [83]	0 [104]	1.4 [67]	1.2 [83]	0 [104]
NAQSRNVGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0 [104]	1.4 [67]	0 [83]	0 [104]
NAQSRNEGTIKDFMNNENSIDAKNAYMTGVDVAE	0 [83]	0 [104]	0 [67]	2.4 [83]	0 [104]
HIQSRNEGTIKDFMNNQNSMEEREAYMPGVDVAQ	0 [83]	0 [104]	0 [67]	1.2 [83]	5.7 [104]
NAQSKNEGTIKDFMNNENSMEEREAYMPGVDVAE	0 [83]	0 [104]	0 [67]	1.2 [83]	0 [104]
NAQSRNEGTIKDFMNNENSIEEREAYMPGVDVAQ	0 [83]	0 [104]	0 [67]	1.2 [83]	0 [104]
NAQSRNEGTIKDFMNNENSMEEREAYIPGVDVAE	0 [83]	0 [104]	0 [67]	1.2 [83]	0 [104]
NAQSRNEGTIKDFMNNENSMEEREAYMTDVDVAQ	0 [83]	0 [104]	0 [67]	1.2 [83]	0 [104]
NAQSRNEGTIKDFMNNENSMEERNAYMTDVDVAE	0 [83]	0 [104]	0 [67]	1.2 [83]	0.9 [104]
NAQSRNEGTIKDFMNNENSMEEREAYMPGVDIPE	0 [83]	0 [104]	0 [67]	0 [83]	2.8 [104]
HAQSRNEGTIKDFMKNENSMEEREAYMPGVDVAQ	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
HAQSRNEGTIKDFMNNENSMEEREAYMPGVDVAE	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
NAHSRNEGTIKDFMNNENTMEEREAYMTGVDVAE	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
NAQSKNEGTIKDFMNNENSMEEREAYMPGVDVAQ	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
NAQSRNEGTIKDFMKNENSMEEREAYMPDVDVAE	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
NAQSRNEGTIKDFMNNENTMEEGEAYMPGVDVAQ	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
NAQSRNEGTIKDFMNNENTMEEREAYMTDVDVAQ	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]
NVHFRNEGTIKDFMNNENSMEEREAYIPGVDVAQ	0 [83]	0 [104]	0 [67]	0 [83]	0.9 [104]

The following were used to define haplotypes: 35 polymorphic sites, including codons 4, 8, 15, 59, 60, 86, 114, 115, 126, 134, 140, 143, 144, 150, 167, 173, 204, 210, 224, 228, 245, 248, 249, 255, 257, 299, 321, 335, 343, 345, 350, 376, 393, 400, 414. The3D7 haplotype was not identified. Frequency is presented as the percentage of sequences that contributed to the haplotype out of the total number of sequences [n]. In grey are zero SNP frequencies.

In *ubp-1*, the following polymorphic codons were observed across all timepoints, E1508*Deletion*, N1518N, KYD-repeat at codon 1520 and KYE-repeat at codon 1526. Only the polymorphism at codon 1520 and 1526 exceeded 10% frequencies across time and each had one stable allele. No SNPs were found to be in LD and the E1528D polymorphism observed in previous studies from Kenya was not identified **Table 4.9**.

Table 4.9. *ubp-1* SNP frequencies

Codon	Nucleotide	Codon [Nucleotide]		Mutant Frequency %[n]				
		Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16
1472	4415	S [C]	Y [A]	0 [83]	0.8 [122]	0.8 [119]	0 [107]	0 [108]
1484	4450	D [G]	Y [T]	0 [83]	0.7 [126]	0 [125]	1.8 [110]	0 [109]
1499	4496	R [G]	H [A]	0 [83]	0.7 [136]	0 [126]	0 [112]	0.9 [110]
1504	4511	E [A]	G [G]	1.1 [84]	0 [136]	0 [127]	0 [112]	0 [110]
1505	4515	K [A]	K [G]	0 [84]	0 [136]	0.7 [127]	0 [112]	0 [110]
1509	4525	E [GAA]	Del	2.3 [83]	3.6 [132]	5.5 [127]	4.4 [112]	4.5 [110]
1514	4540	KNEx2 [AAAAATGAAAAAAACGAA]	2xKNE [AAAAATGAAAAAAATGAA]	0 [86]	9.4 [137]	5.5 [128]	11 [112]	5.4 [112]
			KNEKND [AAAAATGAAAAAAACGAC]	0 [86]		0.7 [128]	0 [112]	0 [112]
			KNE [AAAAATGAA]	0 [86]	1.4 [137]	1.5 [128]	0 [112]	0 [112]
			KNE [AAAAACGAA]	0 [86]		0 [127]	0.8 [112]	0.9 [112]
1515	4545	N [T]	N [C]	0 [86]	0 [137]	0 [127]	0.8 [1]	0 [111]
1518	4554	N [C]	N [T]	8.1 [135]	9.6 [13]	5.6 [123]	11 [13]	5.4 [111]
1519	4557	Y [A]	Y [C]	0 [87	0 [135]	0.8 [125]	0 [111]	0 [110]
1520	4558	KYDx2 [AAATATGACx2]	KYDx4 [AAATATGACx4]	0 [87	11 [137]	3.1 [/127]	0.8 [112]	0 [112]
			KYDx3 [AAATATGACx3]	8.0 [87]	0 [137]	7.0 [127]	8.0 [112]	11 [112]
			KYGKYD [AAATATGGCAAATATGAC]	0 [87	0.7 [137]]	0 [127]	0 [112]	0 [112]
			KYD [AAATATGAC]	10 [87]	13 [137]	18 [127]	16 [112]	12 [112]
1522	4565	D [A]	G [G]	0 [87]	0.7 [137]	0 [127]	0 [112]	0 [111]
1526	4576	KYEx2 [AAATATGAAx2]	KYEKYD[KYEx2] [AAATATGAAAAATATGAC(AAATATGAAx2)]	0 [87]	0 [135]	0.7 [127]	0 [112]	0 [112]
			KYEx3 [AAATATGAAx3]	11 [87]	14 [135]	12 [127]	10 [112]	6.3 [112]
			KYEx3 [AAATACGAA(AAATATGAAx2)]	0 [87]	0.7 [135]	0 [127]	0 [112]	0 [112]
			KYEKYDKYE [AAATATGAAAAATATGACAAATATGAA]	0 [87]	0 [135]	1.5 [127]	0 [112]	0 [112]
			KYEKYV [AAATATGAAAAATATGTA]	0 [87]	0.7 [135]	0 [127]	0 [112]	0 [112]
			KYE [AAATATGAA]	17 [15]	11 [135]	11 [14]	12 [14]	17 [112]
1527	4581	Y [T]	Y [C]	0 [87]	0.7 [133]	0 [127]	0 [112]	0 [111]
1531	4592	E [A]	V [T]	0 [72]	0.8 [122]	0 [113]	0 [98]	0 [92]
1532	4594	KYD [AAATATGAT]	KYDx2 [(2xAAATATGAT)]	1.1 [85]	0 [135]	0 [126]	0 [111]	0 [111]
1575	4724	K [A]	R [G]	0 [86]	0 [135]	0.7 [126]	0 [105]	0 [108]

The number of samples successfully genotyped per timepoint include: 85 in 1995/96, 137 in 1998/99, 128 in 2005/06, 112 in 2012/13 and 85 in 2015/16. No sequences with mixed bases were identified. Frequency is presented as the percentage of sequences that carried a mutation out of the total number of sequences that had data for that locus [n]. In grey are zero SNP frequencies.

Of the 23 *ubp-1* haplotypes, the 3D7 haplotype dominated throughout the sampling period and no significant temporal trends were observed (**Table 4.10**).

Table 4.10. *ubp-1* haplotype frequencies

Haplotype	Frequency % [n]				
	1995/96	1998/99	2005/06	2012/13	2015/16
SDREE[KNEx2][KYDx2]D[KYEx2]E[KYDx1]*	66.6 [66]	62.2 [106]	58.2 [103]	62.9 [89]	70.9 [86]
SDREE[KNEx2][KYDx1]D[KYEx2]E[KYDx1]	10.6 [66]	11.3 [106]	16.5 [103]	14.6 [89]	12.7 [86]
SDREE[KNEx2][KYDx2]D[KYEx3]E[KYDx1]	9.0 [66]	10.3 [106]	8.7 [103]	10.1 [89]	4.6 [86]
SDREE[KNEx2][KYDx3]D[KYEx2]E[KYDx1]	4.5 [66]	2.8 [106]	3.8 [103]	2.2 [89]	2.3 [86]
SDRE-[KNEx2][KYDx2]D[KYEx3]E[KYDx1]	1.5 [66]	1.8 [106]	0.9 [103]	0 [89]	0 [86]
SDRE-[KNEx2][KYDx3]D[KYEx2]E[KYDx1]	1.5 [66]	0 [106]	0 [103]	1.1 [89]	3.4 [86]
SDREE[KNEx2][KYDx1]D[KYEx3]E[KYDx1]	1.5 [66]	0.9 [106]	0 [103]	1.1 [89]	1.1 [86]
SDREE[KNEx2][KYDx2]D[KYEx2]E[KYDx2]	1.5 [66]	0 [106]	0 [103]	0 [89]	0 [86]
SDREE[KNEx2][KYDx3]D[KYEx3]E[KYDx1]	1.5 [66]	0.9 [106]	1.9 [103]	0 [89]	0 [86]
SDRGE[KNEx2][KYDx2]D[KYEx2]E[KYDx1]	1.5 [66]	0 [106]	0 [103]	0 [89]	0 [86]
SDREE[KNEx2][KYDx2]D[KYEx1]E[KYDx1]	0 [66]	1.8 [106]	0 [103]	0 [89]	0 [86]
SDREE[KNEx2][KYDx4]D[KYEx2]E[KYDx1]	0 [66]	1.8 [106]	1.9 [103]	0 [89]	0 [86]
SDRE-[KNEx2][KYDx1]D[KYEx2]E[KYDx1]	0 [66]	0.9 [106]	1.9 [103]	3.3 [89]	0 [86]
SDREE[KNEx1][KYDx3]D[KYEx2]E[KYDx1]	0 [66]	0.9 [106]	0 [103]	0 [89]	1.1 [86]
SDREE[KNEx2][KYDx2]D[KYEKYV]V[KYDx1]	0 [66]	0.9 [106]	0 [103]	0 [89]	0 [86]
SDREE[KNEx2][KYGKYD]G[KYEx3]E[KYDx1]	0 [66]	0.9 [106]	0 [103]	0 [89]	0 [86]
SYREE[KNEx2][KYDx2]D[KYEx2]E[KYDx1]	0 [66]	0.9 [106]	0 [103]	2.2 [89]	0 [86]
YDREE[KNEx1][KYDx2]D[KYEx3]E[KYDx1]	0 [66]	0.9 [106]	0.9 [103]	0 [89]	0 [86]
SDRE-[KNEx2][KYDx4]D[KYEx2]E[KYDx1]	0 [66]	0 [106]	1.9 [103]	0 [89]	0 [86]
SDREE[KNEx2][KYDx2]D[KYEKYDKYE]E[KYDx1]	0 [66]	0 [106]	1.9 [103]	0 [89]	0 [86]
SDREE[KNEx1][KYDx2]D[KYEx3]E[KYDx1]	0 [66]	0 [106]	0.9 [103]	1.1 [89]	0 [86]
SDRE-[KNEx2][KYDx2]D[KYEx2]E[KYDx1]	0 [66]	0 [106]	0 [103]	1.1 [89]	2.3 [86]
SDHEE[KNEx2][KYDx2]D[KYEx3]E[KYDx1]	0 [66]	0 [106]	0 [103]	0 [89]	1.1 [86]

The following were used to define haplotypes: 12 polymorphic sites, including codons 1472, 1484, 1499, 1504, 1508, 1514, 1520, 1522, 1526, 1531, 1532 and 1575. The haplotypes marked with an * indicate the 3D7 haplotype. "-" in some of the haplotypes indicates a deletion. Frequency is presented as the percentage of sequences that contributed to the haplotype out of the total number of sequences [n]. In grey are zero SNP frequencies.

4.3.3 Genetic markers associated with CQ, SP and lumefantrine resistance

Three polymorphic codons (M74I, N75E and K76T) were identified in the *crt* gene in 2015/16 and 2017/18 and all were found to be in high LD (*D'* = 1). Mutant alleles at codons 74I, 75E and 76T dominated during the pre-ACT period and there was a distinct shift to the wild-type alleles (M74, N75 and K76) in the post-ACT period, almost reaching fixation (99%) ($\chi 2$ = 181, p-value <0.001, **Table 4.11**). There was a significant decline in the frequency of the CQ-resistant (CQR) haplotype (CVIET) over time reaching 1% in 2017/18 and a sharp increase in

the frequency of the CQ-sensitive (CQS) haplotype (CVMNK) from 7% in 1998/99 to 99% in 2017/18 ($\chi 2$ = 181, p-value <0.001, **Figure 4.5A, Table 4.12**).

Five polymorphic codons were identified in *mdr1* in 2015/16 and 2017/18 including N86Y, G102G, G182G, F184Y and D1246Y. *mdr1* codons 86Y and 1246Y also showed a distinct shift from mutant alleles in the pre-ACT period (~40-70%), to wild-type alleles N86 and D1246, nearly approaching fixation (99%) ($\chi 2$ = 103, p-value <0.001 and $\chi 2$ = 85, p-value <0.001, respectively) post-ACT introduction. In contrast, the mutant 184F allele increased in frequency during the post-ACT period (33-54%, $\chi 2$ = 15.8, p-value <0.001, **Table 4.11**). Codons N86Y and F186Y as well as N86Y and D1246Y were found to be in high and low LD, respectively (*D'* = 0.9 and *D'* = -1, respectively). There were notable changes in frequency of both the 3D7 haplotype (NYD) and the mutant NFD haplotype. NYD increased sharply to 64% in 2012/13, decreased to 40% in 2015/16 and later rose to 55% in 2017/18 to become the dominant haplotype. The mutant haplotype NFD followed an opposite pattern rising to 55% in 2015/16 and later decreasing to 41% in 2017/18. The triple mutant haplotype YYY was no longer detected in the population post-ACT (**Figure 4.5B, Table 4.12**) and there was no significant temporal trend observed for the haplotype frequencies.

S436A, A437G, K540E and A581G *dhps* polymorphic codons were identified across all time points and the I431V mutation was not identified. The mutant (437G and 540E) alleles dominated in the post-ACT period (68-91%, $\chi 2$ = 82.7, p<0.001 and $\chi 2$ = 153, p<0.001, respectively). Moreover, codons A437G and K540E were found to be in high LD (*D'* = 1). The SP sensitive haplotype (SAKA) decreased in frequency over time and was no longer detectable in 2015/16 (**Table 4.11**), while the single mutant haplotype (SGKA) decreased gradually to 9% in 2015/16 and the double mutant haplotype (SGEA) rose in frequency from 10% to 85% in 2015/16. (**Figure 4.5C, Table 4.12**). There was a significant temporal trend between the two dominant haplotypes, SAKA and SGEA ($\chi 2$ = 91, p-value <0.001).

Table 4.11. *crt*, *mdr1* and *dhps* SNP frequencies

Gene	Codon	NT	Codon [NT]		Mutant Frequency % [n]					
			Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16	2017/18
crt	74	222	M [G]	I [T]	62.03 [80]	93.2 [103]	50.98 [102]	18.29 [82]	3.16 [94]	1.1 [92]
	75	223	N [A]	E [G]	62.03 [80]	93.2 [103]	50.98 [102]	18.29 [82]	3.16 [94]	1.1 [92]

	75	225	N [T]	E [A]	62.03 [80]	93.2 [103]	50.98 [102]	18.29 [82]	3.16 [94]	1.1 [92]
	76	227	K [A]	T [C]	62.03 [80]	93.2 [103]	50.98 [102]	18.29 [82]	3.19 [94]	1.1 [92]
	86	256	N [A]	Y [T]	57.14 [57]	72.84 [81]	57.45 [47]	2.08 [48]	2.8 [107]	1.22 [82]
	102	306	G [T]	G [C]	ND	ND	ND	ND	1.9 [105]	7.41 [81]
mdr1	182	546	G [T]	G [G]	ND	ND	ND	ND	2.56 [78]	0 [89]
	184	551	Y [A]	F [T]	30.56 [57]	13.58 [81]	29.79 [47]	33.33 [48]	54.43 [79]	42.5 [80]
	1246	3736	D [G]	Y [T]	37.5 [57]	64.2 [81]	38.3 [47]	0 [48]	2.8 [107]	1.14 [88]
	436	1306	S [T]	A [G]	0 [85]	2.42 [124]	0 [80]	0 [21]	0 [24]	NA
dhps	437	1310	A [C]	G [G]	32.94 [85]	37.9 [124]	86.49 [74]	78.57 [28]	0 [26]	NA
	540	1618	K [A]	E [G]	10.42 [196]	28.12 [128]	83.17 [101]	68.29 [41]	91.3 [45]	NA
	581	1742	A [C]	G [G]	0 [93]	0 [120]	0 [89]	0 [54]	3.03 [32]	NA

The number samples successfully genotyped per timepoint include: *crt* – 103 in 2015/16 and 91 in 2017/18 and no sequences with mixed bases identified. *mdr1* – 130 in 2015/16 and 88 in 2017/18 and no sequences with mixed bases identified. *dhps* – 99 in 1995/96, 137 in 1998/99, 130 in 2005/06, 80 in 2012/13 and 72 in 2015/16 with two sequences having mixed bases in 1998/99. For *dhps* 2017/18 and *crt* (codons 102 and 182 in 1995/96, 1998/99 and 2005/06, respectively) no data (NA) was available because they were not genotyped. *NT* – Nucleotide, *(*)* one-sided, 97.5% confidence interval. In grey are zero SNP frequencies.

Table 4.12. *crt*, *mdr1* and *dhps* haplotype frequencies

Gene	Haplotype	Frequency % [n]					
		1995/96	**1998/99**	**2005/06**	**2012/13**	**2015/16**	**2017/18**
crt	CVMNK*	37.9 [79]	6.8 [103]	49.0 [102]	81.7 [82]	96.7 [91]	98.8 [90]
	CVIET	62.0 [79]	93.2 [103]	50.9 [102]	18.2 [82]	3.3 [91]	1.1 [90]
mdr1	NYD*	14.2 [56]	12.3 [81]	12.7 [47]	64.5 [48]	40 [60]	55.7 [70]
	NFD	28.5 [56]	7.4 [81]	12.7 [47]	33.3 [48]	55 [60]	41.4 [70]
	YFY	1.7 [56]	0 [81]	0 [47]	0 [48]	0 [60]	0 [70]
	YYD	19.6 [56]	16.0 [81]	31.9 [47]	2.0 [48]	3.3 [60]	1.4 [70]
	YYY	35.7 [56]	56.7 [81]	21.2 [47]	0 [48]	0 [60]	0 [70]
	NFY	0 [56]	6.1 [81]	12.7 [47]	0 [48]	0 [60]	1.4 [70]
	NYY	0 [56]	1.2 [81]	4.2 [47]	0 [48]	1.6 [60]	0 [70]
	YFD	0 [56]	0 [81]	4.2 [47]	0 [48]	0 [60]	0 [70]
dhps	SGEA*	10.7 [84]	31.6 [115]	84.2 [57]	76 [25]	85.7 [21]	NA
	SAKA	66.6 [84]	59.8 [115]	14.0 [57]	16 [25]	0 [21]	NA
	SGKA	22.6 [84]	5.9 [115]	1.7 [57]	8 [25]	9.5 [21]	NA
	AAKA	0 [84]	2.5 [115]	0 [57]	0 [25]	0 [21]	NA
	SGEG	0 [84]	0 [115]	0 [57]	0 [25]	4.7 [21]	NA

The following were used to define haplotypes: *crt* – 3 polymorphic sites, including codons 74, 75 and 76. *Mdr1* – 3 polymorphic sites, including codons 86, 184 and 1246 and *dhps* – 4 polymorphic sites, including codons 436, 437, 540 and 581. The haplotypes marked with an * indicate the 3D7 haplotype. For *dhps* , there was no data for the 2017/18 timepoint as it was not genotyped. Frequency is presented as the percentage of sequences that contributed to the haplotype out of the total number of sequences [n]. In grey are zero SNP frequencies.

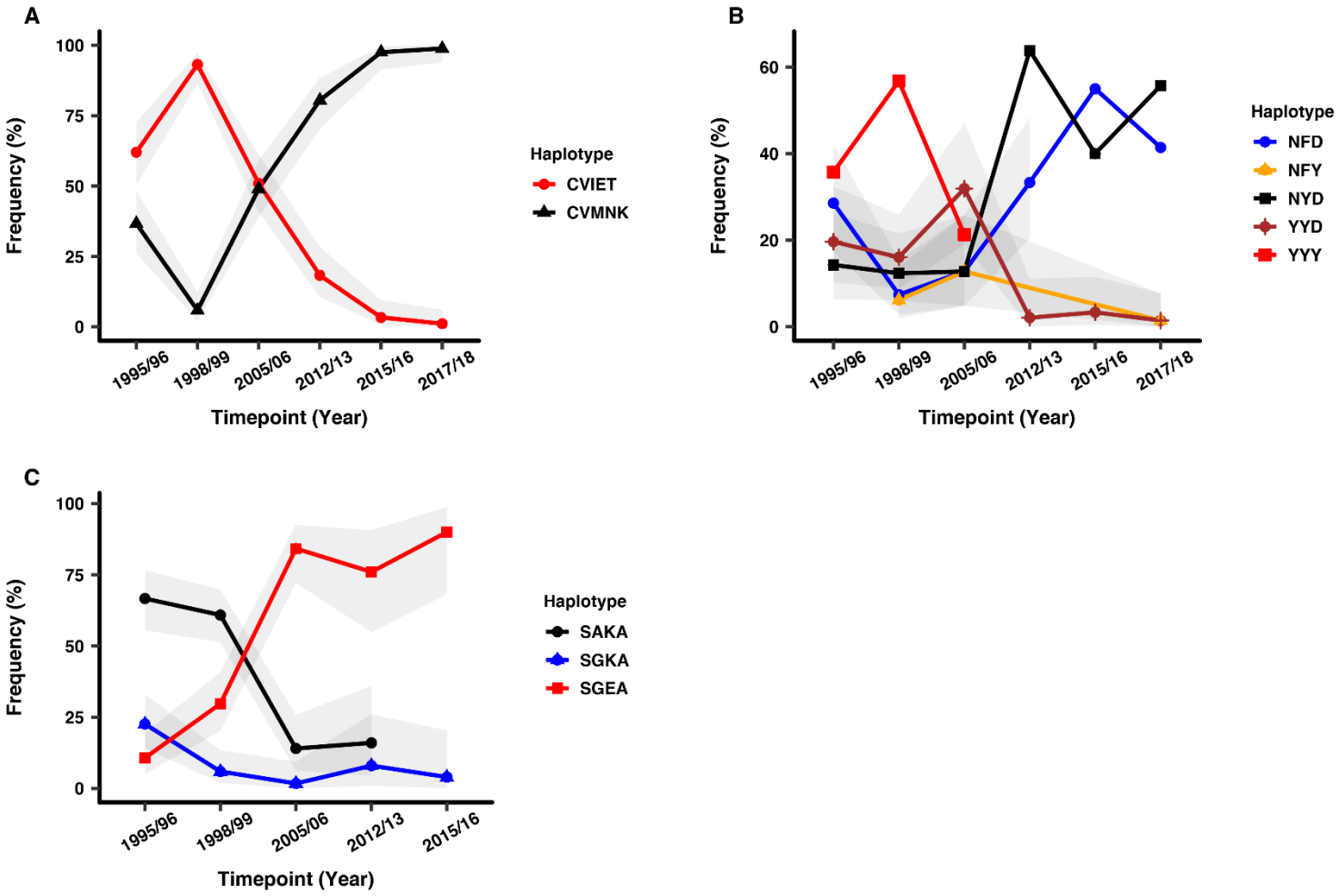


Figure 4.5. *crt*, *mdr1*, *dhps* haplotypes frequencies over time. (A) The *crt* sensitive haplotype (CVMNK) decreased from 1995/96 to 1998/99 and increased onwards to almost fixation in 2017/18 while the *crt* resistant haplotype (CVIET) followed an opposite pattern. (B) The 3D7 *mdr1* haplotype NYD had was the least prevalent in comparison to the mutant haplotypes, NFD, YYD and YYY pre-ACT introduction. The triple mutant YYY was undetectable after 2005/06 while the 3D7 NYD and mutant NFD haplotypes started to increase in the population after 2005/06. Mutants YYD and NFY haplotypes decreased to almost zero in 2017/18. (C) The SP sensitive haplotype (SAKA) was on a decline from 1995/96 and was undetectable in the population after 2012/13. The SP resistant double mutant haplotype (SGEA) was on the increase from 1995/96 and reached >80% frequency in 2015/16. The single mutant haplotype (SGKA) was the least prevalent throughout the sampling period. In grey are the 95% confidence intervals.

In *nfs*, only codon K65Q was found to have a significant, albeit marginal, trend pre- and post-ACT introduction ($\chi 2 = 4.4$, p-value = 0.04). It was also in high LD ($D' = 1$) with codons S62N and E67G. The 2005/2006 period appears to be the point at which the mutant allele 65Q and

wild-type 65K diverge in frequency in opposite directions with the latter decreasing and the former increasing in frequency (**Figure 4.6** and **Table 4.13**).

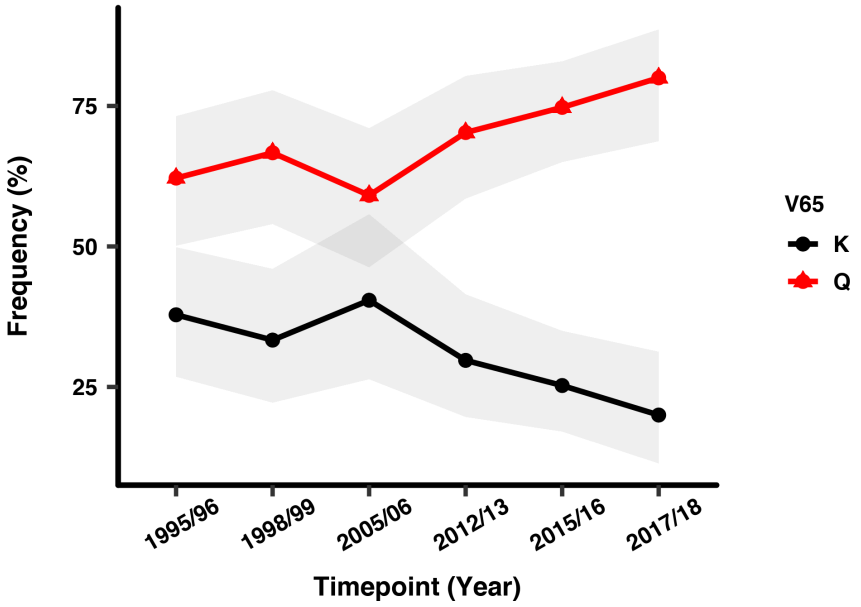


Figure 4.6. *nfs* codon K65Q frequencies over time. The two *nfs* K65Q alleles appear to have stable frequencies from 1995/96 to 2005/06 but the frequency of K65 starts to drop after 2005/06 while that of 65K starts to increase after 2005/06. In grey are the 95% confidence intervals.

Table 4.13. *nfs* SNP frequencies

Codon	Nucleotide	Codon [Nucleotide]		Mutant Frequency % [n]					
		Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16	2017/18
44	131	S [C]	C [G]	0 [73]	0 [64]	2.0 [48]	4.4 [68]	6.1 [97]	4.2 [70]
62	185	S [G]	N [A]	37 [74]	33 [64]	36 [47]	28 [73]	25 [99]	20 [71]
65	193	K [A]	Q [C]	62 [74]	66 [64]	40 [47]	70 [74]	74 [99]	80 [70]
67	200	E [A]	G [G]	62 [74]	66 [64]	40 [47]	70 [74]	74 [79]	80 [70]
	201	E [A]	G [T]	62 [74]	66 [64]	40 [47]	70 [74]	74 [79]	80 [70]
110	328	Q [C]	E [G]	5.4 [74]	10 [65]	8 [50]	1.2 [78]	16 [79]	5.7 [70]
116	347	S [G]	I [T]	0 [74]	1.5 [66]	0 [51]	0 [79]	0 [100]	98 [70]
119	355	G [G]	C [T]	0 [74]	3.0 [67]	0 [51]	1.2 [79]	0 [100]	0 [70]
120	359	S [G]	I [T]	14 [74]	27 [66]	37 [51]	30 [79]	41 [100]	32 [70]
126	376	P [C]	S [T]	0 [74]	0 [65]	1.9 [51]	0 [79]	0 [100]	0 [70]
130	389	E [A]	G [G]	0 [74]	0 [65]	0 [51]	1.2 [79]	0 [100]	0 [70]
	390	E [G]	D [C]	13 [74]	18 [65]	15 [51]	18 [79]	25 [25]	22 [70]
188	562	E [G]	K [A]	0 [24]	0 [57]	0 [44]	0 [60]	0 [92]	5.9 [67]
274	822	G [T]	G [C]	0 [76]	0 [66]	0 [57]	0 [82]	0 [100]	1.4 [70]
283	847	N [A]	Y [T]	0 [76]	0 [66]	0 [58]	0 [82]	0 [100]	1.4 [70]
288	864	N [T]	N [C]	0 [76]	0 [66]	1.7 [58]	0 [82]	0 [100]	0 [70]
308	922	E [G]	Q [C]	0 [76]	0 [66]	0 [57]	0 [82]	0 [100]	1.4 [70]
334	1002	I [C]	I [A]	1.3 [75]	0 [66]	0 [56]	0 [82]	0 [100]	0 [70]

338	1013	K [A]	R [G]	0 [75]	0 [66]	0 [56]	0 [82]	0 [100]	1.4 [70]
379	1137	G [C]	G [T]	0 [74]	0 [66]	0 [54]	0 [82]	0 [100]	1.4 [70]
409	1225	N [A]	Y [T]	0 [74]	0 [66]	0 [53]	0 [82]	0 [102]	2.9 [69]

The number of samples successfully genotyped per timepoint include: 81 in 1995/96, 71 in 1998/99, 73 in 2005/06, 86 in 2012/13, 114 in 2015/16 and 77 in 2017/18. There were a total of 4, 3, 3, 4, 11 and 7 sequences with mixed bases in 1995/96, 1998/99, 2005/06, 2012/13, 2015/16 and 2017/18, respectively. Frequency is presented as the percentage of sequences that carried a mutation out of the total number of sequences that had data for that locus [n]. In grey are zero SNP frequencies.

The 3D7 *nfs* haplotype was the seventh most dominant of all the 24 haplotypes and there were no significant temporal trends in haplotype frequencies (**Table 4.14**).

Table 4.14. *nfs* haplotype frequencies

Haplotype	Frequency % [n]					
	1995/96	1998/99	2005/06	2012/13	2015/16	2017/18
SNQGQSGIPE	27.2 [22]	7.1 [56]	17.0 [41]	15.2 [59]	9.7 [94]	12.8 [70]
SSKEQSGSPD	22.7 [22]	3.5 [56]	2.4 [41]	5.0 [59]	4.3 [94]	2.8 [70]
SNQGQSGSPD	13.6 [22]	10.7 [56]	0 [41]	6.7 [59]	3.2 [94]	8.5 [70]
SNQGQSGSPE	13.6 [22]	37.5 [56]	31.7 [41]	37.2 [59]	28.2 [94]	41.4 [70]
SNQGESGIPE	9.0 [22]	7.1 [56]	4.8 [41]	0 [59]	6.5 [94]	1.4 [70]
SNQGQSGIPD	9.0 [22]	3.5 [56]	2.4 [41]	8.4 [59]	13.0 [94]	5.7 [70]
SSKEQSGSPE*	4.5 [22]	10.7 [56]	12.2 [41]	5.0 [59]	8.7 [94]	7.1 [70]
SSKEQSGIPE	0 [22]	7.1 [56]	7.3 [41]	8.4 [59]	4.3 [94]	5.7 [70]
SNQGESGSPE	0 [22]	3.5 [56]	2.4 [41]	1.6 [59]	6.5 [94]	4.2 [70]
SSKEQSCIPE	0 [22]	3.5 [56]	0 [41]	1.6 [59]	0 [94]	0 [70]
SSKEESGSPE	0 [22]	1.7 [56]	0 [41]	0 [59]	0 [94]	0 [70]
SSKEQIGIPD	0 [22]	1.7 [56]	0 [41]	0 [59]	0 [94]	0 [70]
SSKEQSGIPD	0 [22]	1.7 [56]	7.3 [41]	3.3 [59]	5.4 [94]	4.2 [70]
CSKEQSGSPD	0 [22]	0 [56]	2.4 [41]	0 [59]	0 [94]	0 [70]
SNKEQSGSPD	0 [22]	0 [56]	2.4 [41]	0 [59]	0 [94]	0 [70]
SNKEQSGSPE	0 [22]	0 [56]	2.4 [41]	0 [59]	0 [94]	0 [70]
SNQGQSGSSE	0 [22]	0 [56]	2.4 [41]	0 [59]	0 [94]	0 [70]
SSKEESGIPE	0 [22]	0 [56]	2.4 [41]	0 [59]	2.1 [94]	0 [70]
CNQGQSGIPD	0 [22]	0 [56]	0 [41]	1.6 [59]	0 [94]	1.4 [70]
CNQGQSGIPE	0 [22]	0 [56]	0 [41]	1.6 [59]	2.1 [94]	1.4 [70]
CNQGQSGSPE	0 [22]	0 [56]	0 [41]	1.6 [59]	4.3 [94]	1.4 [70]
SNQGQSGSPG	0 [22]	0 [56]	0 [41]	1.6 [59]	0 [94]	0 [70]
SNQGESGIPD	0 [22]	0 [56]	0 [41]	0 [59]	1.0 [94]	0 [70]
SNQGQNGSPE	0 [22]	0 [56]	0 [41]	0 [59]	0 [94]	1.4 [70]

The following were used to define haplotypes: 10 polymorphic sites, including codons 44, 62, 65, 67, 110, 116, 119, 120, 126 and 130. The haplotypes marked with an * indicate the 3D7 haplotype. Frequency is presented as the percentage of sequences that contributed to the haplotype out of the total number of sequences [n]. In grey are zero SNP frequencies.

4.3.4 *serine-tRNA ligase, putative*

serine tRNA ligase, a marker not associated with resistance or dug selection, had only one polymorphic codon observed across all time points (L84V), with frequencies ranging between 2- 5% while the rest were rare and no SNPs were found to be in LD (<5%, **Table 4.15**).

Table 4.15. *serine-tRNA ligase* SNP frequencies

Codon	Nucleotide	Codon [Nucleotide]		Mutant Frequency %[n]				
		Wildtype	Mutant	1995/96	1998/99	2005/06	2012/13	2015/16
31	92	N [A]	S [G]	0 [95]	0 [120]	0 [105]	0.8 [114]	3 [127]
71	211	D [G]	N [A]	2 [105]	0 [133]	0 [121]	0 [121]	0 [142]
84	250	L [T]	V [G]	3 [106]	3 [133]	5 [120]	4 [122]	2 [142]
86	258	I [A]	M [G]	0 [106]	0 [133]	0 [121]	0 [122]	1 [142]
88	264	E [A]	D [T]	0 [106]	6 [133]	0.8 [121]	1 [122]	2 [142]
94	280	Q [C]	E [G]	0 [107]	0 [133]	0 [121]	0.8 [122]	2 [142]
133	399	T [A]	T [G]	0 [106]	2 [132]	0.8 [121]	0 [120]	0.7 [142]
140	418	L [C]	F [T]	0 [105]	0.7 [132]	0 [118]	0 [119]	0 [142]
155	463	V [G]	L [T]	0 [105]	0 [133]	0 [120]	1 [118]	0 [142]
175	524	I [T]	K [A]	0 [95]	0 [128]	0.9 [108]	0 [114]	0 [139]
199	596	A [C]	V [T]	0 [91]	0 [116]	0 [190]	0 [113]	0.7 [134]
221	662	A [C]	G [G]	0 [94]	0.8 [124]	0.9 [111]	1 [120]	0.7 [138]
267	799	T [A]	P [C]	0 [93]	0 [126]	0 [113]	5 [118]	1 [136]
284	851	A [C]	V [T]	0 [95]	0.7 [128]	0 [110]	0 [117]	0 [136]
318	952	E [G]	K [A]	0 [94]	0.7 [130]	0 [107]	0 [115]	0 [139]
372	1116	F [F]	F [T]	1 [89]	0 [127]	0 [107]	0 [115]	0 [139]
417	1249	S [T]	A [G]	0 [92]	1 [126]	0 [105]	0 [117]	0.7 [139]
430	1290	N [T]	N [C]	0 [92]	1 [129]	0 [107]	0 [120]	0 [139]
528	1584	Y [C]	Y [T]	0 [99]	0 [126]	0 [110]	1 [101]	0 [139]

The number of samples successfully genotyped per timepoint include: 118 in 1995/96, 136 in 1998/99, 130 in 2005/06, 123 in 2012/13 and 118 in 2015/16. No sequences with mixed bases were identified. Frequency is presented as the percentage of sequences that carried a mutation out of the total number of sequences that had data for that locus [n]. In grey are zero SNP frequencies.

A total of 15 *serine-tRNAligase* haplotypes were observed and only the 3D7 haplotype occurred across all time points and with frequencies >80% with the rest of the haplotypes being rare (<5%) (**Table 4.16**).

Table 4.16. *serine-tRNAligase* haplotype frequencies

Haplotype	Frequency % [n]				
	1995/96	1998/99	2005/06	2012/13	2015/16
NDLIEQLVAATAES*	93.2 [74]	85.2 [102]	95.8 [72]	86.1 [101]	84.6 [117]
NNLIEQLVAATAES	4.0 [74]	0 [102]	0 [72]	0 [101]	0 [117]
NDVIEQLVAATAES	2.7 [74]	3.9 [102]	4.1 [72]	1.9 [101]	2.5 [117]

NDLIDQLVAATAES	0 [74]	4.9 [102]	0 [72]	0.9 [101]	3.4 [117]
NDLIEQLVAATAEA	0 [74]	1.9 [102]	0 [72]	0 [101]	0.8 [117]
NDLIDQLVAATAKS	0 [74]	0.9 [102]	0 [72]	0 [101]	0 [117]
NDLIDQLVAATVES	0 [74]	0.9 [102]	0 [72]	0 [101]	0 [117]
NDLIEQFVAATAES	0 [74]	0.9 [102]	0 [72]	0 [101]	0 [117]
NDLIEQLVAGTAES	0 [74]	0.9 [102]	0 [72]	1.9 [101]	0.8 [117]
NDLIEQLVAAPAES	0 [74]	0 [102]	0 [72]	5.9 [101]	1.7 [117]
NDLIEQLLAATAES	0 [74]	0 [102]	0 [72]	1.9 [101]	0 [117]
NDLIEELVAATAES	0 [74]	0 [102]	0 [72]	0.9 [101]	0.8 [117]
SDLIEQLVAATAES	0 [74]	0 [102]	0 [72]	0 [101]	3.4 [117]
NDLIEQLVVATAES	0 [74]	0 [102]	0 [72]	0 [101]	0.8 [117]
NDLMEQLVAATAES	0 [74]	0 [102]	0 [72]	0 [101]	0.8 [117]

The following were used to define haplotypes: 14 polymorphic sites, including codons 31, 71, 84, 86, 88, 94, 140, 155, 199, 221, 267, 284, 318 and 417. The haplotypes marked with an * indicate the 3D7 haplotype. Frequency is presented as the percentage of sequences that contributed to the haplotype out of the total number of sequences [n]. In grey are zero SNP frequencies.

4.4 Discussion

The withdrawal of CQ in Kenya (Shretta *et al.* 2000) has resulted in the rapid decline in CQR alleles, namely *crt*-76T and *mdr1*-86Y and 1246Y, to 1% and the complete absence of the triple mutant *mdr1*-YYY, that modulates sensitivity to CQ and AQ (Reed *et al.* 2000; Humphreys *et al.* 2007), in the Kilifi population. Additionally, there has been a near complete reversion to the wild-type alleles *crt*-K76 (99%), *mdr1*-N86 (97%) and D1246 (99%). This is contrary to the neutral marker evaluated*, serine tRNA ligase*, which showed a dominance of the wild-type allele (>90%) and haplotype (NDLIEQLVAATAES, >80%) across the entire sampling period with no significant temporal differences. It is unclear whether this trend will continue so that the frequency of the *crt*-76T mutant allele approaches zero, or if this allele will persist at a new, very low steady state frequency. This observed trend supports a previous study that predicted it would take an additional 13 years to restore CQ clinical efficacy in Kilifi based on sampling between 1993 to 2006 (Mwai *et al.* 2009b). This near complete reversion to *crt*-CVMNK in Kilifi might be taken to suggest that after 2019, CQ could be re-introduced as a treatment option for malaria. Nationally, a decline in CQR has previously been observed in Kwale on the South Coast of Kenya, *crt*-K76 was on the rise at 36% in 2008 from 22% in 1999 (Mang'era *et al.* 2012), Kisumu, Kisii and Kericho in the Western region of Kenya (Eyase *et al.* 2013; Achieng *et al.* 2015). In Kisumu where malaria transmission is holoendemic, between the years 2003

and 2014, *crt*-76T dropped from 86% to 2%, *mdr1*-86Y from 92% to 1% and *mdr1*-1246Y from 67% to 6%. Therefore, a national re-roll out of CQ in the next few years is a possibility, such as in a combination therapy, however careful consideration needs to be taken to determine the best modality for its re-introduction. As for *mdr1*, the NFD haplotype appears to oscillate in frequency with the NYD (wild type) haplotype and both have gone on to be the dominant *mdr1* haplotypes post-ACT introduction. The NYD haplotype is likely to have risen in frequency due to the absence of CQ, restoring the wild type parasite population while the increase in the NFD haplotype frequency may be attributable to AL pressure (Humphreys et al., 2007). The *mdr1*-G102G and G182G identified in 2015/16 and 2017/18 were not genotyped in an earlier study (Okombo *et al.* 2014) hence their associations with changing anti-malarial drug policy could not be described.

The introduction of SP in 1998 led to a rapid increase in the frequency of the *dhps* mutations (437G/540E) associated with resistance to sulfadoxine. This upsurge is in accordance with the previously observed increase in *dhfr* mutations (51I/59R/108N) associated with pyrimethamine resistance in the same population (Okombo *et al.* 2014) and the high frequency of SP resistance markers in Kenya may be attributable to the continued distribution of SP for malaria case management in the private sector (Musuva *et al.* 2017). Even though SP is no longer the first line treatment for malaria, similar observations have been made in others regions of SSA (Okell, Griffin and Roper 2017). Still, SP maintains its utility in IPTp but the loss of IPTp efficacy has been noted when the prevalence of sextuple-mutant parasites (*dhfr* 51I/59R/108N and *dhps* 437G/540E/581G) exceeds 37% (van Eijk *et al.* 2019), as seen in some SSA settings. Consequently, the *dhps*-581G mutation needs to be monitored given that it first occurs in our population in 2015/16 at 3%. The *dhps* I431V allele was not detected and its distribution appears to be restricted to West Africa (Oguike *et al.* 2016).

The decline in the *nfs*-K65 wild-type allele in the Kilifi population draws attention to the potential causes of its decline since the introduction of AL in 2004. Recent findings from The Gambia (West Africa) implicated lumefantrine as a selective pressure, showed an increase in *nfs*-K65 allele frequencies over time and described its association with a higher lumefantrine IC_{50} compared to the mutant 65Q allele (Amambua-Ngwa *et al.* 2018). It therefore appears

that there is an opposite trend of the K65 allele in this East African population. The Gambian parasites have shown increasing tolerance to lumefantrine in a study conducted between 2013 and 2015 (Amambua-Ngwa *et al.* 2017). However, drug trials with AL in Western Africa, including The Gambia, show that AL is still highly efficacious (Dieye *et al.* 2016). This is despite sporadic reports of AL treatment failure in Swedish and UK travellers to SSA that may be an early indication of falling AL efficacy (Sondén *et al.* 2017; Sutherland *et al.* 2017). It's worth noting that Sondén *et al.* (2017) attributed the treatment failures to lower lumefantrine concentrations during the initial AL treatment. On the other hand, Sutherland *et al.* (2017) mentioned that the patients with treatment failures were not observed to ensure adherence to AL regimen nor did they measure lumefantrine blood levels at day 7, to rule out any problems with malabsorption, as this may have resulted in failure to clear parasites. However, Of equal interest is the increase in the 65Q allele in the Kilifi population compared to the apparent stable frequency in the Gambia. Perhaps these discordant observations are due to differences in allele frequencies of other loci such as *pfmdr1* between East and West Africa (Okombo *et al.* 2013), or the differences in drug policy (such as the extensive use of amodiaquine in West Africa compared to East Africa) (Okell *et al.* 2018) or perhaps the inverse relationship between LM and CQ resistance (Mwai *et al.* 2009a).

The common *k13* A578S mutation that has been found to be prevalent in Africa (Kamau *et al.* 2015; MalariaGEN Plasmodium falciparum Community Project 2016; Ménard *et al.* 2016) was also observed in our population at lower frequencies <2% across all time points. In contrast, the K189T mutation was the only SNP with a stable, >10%, frequency across all time points and it appears to be prevalent globally (MalariaGEN Plasmodium falciparum Community Project 2016). Notably, a recent study confirmed that the A578S mutation does not confer artemisinin resistance by carrying out genome editing to confirm its phenotypic effect (Ménard *et al.* 2016), consistent with limited field data (Muwanguzi *et al.* 2016). This observation was comparable to *serine-tRNA ligase*, the neutral marker that was not under drug pressure. Moreover, similar to parasites from SSA (MalariaGEN Plasmodium falciparum Community Project 2016), the N-terminal region of *k13* had higher Ka/Ks ratios than the C-terminal region since a large majority of the polymorphisms were identified in this region of

the gene. This is not the case in SE Asia where the C-terminal region that harbours artemisinin resistance SNPs shows higher Ka/Ks ratios, which is most likely the result of artemisinin selection pressure. Similar to previous studies (MalariaGEN Plasmodium falciparum Community Project 2016), the artemisinin resistance predisposing mutations were not identified, apart from a high frequency SNP (I492V) in *mdr2* that did not show evidence of selection over time. This mutation (I492V) has also been identified at 100% frequency (n = 38) in Suriname (Chenet *et al.* 2017).

The distinct patterns of selection observed with *crt*, *mdr1* and *dhps* were not seen in the other genes evaluated including *ap2-mu*, *falcipain-2a*, *nfs* and *ubp-1*. Notably, the *ap2-mu*-160N mutation was observed across all time points at stable frequencies, showing no evidence of selection. However, the *ap2-mu*-S160N mutation has previously been found to be selected for after ACT treatment although it was not associated with treatment failure in a Western Kenya population (Henriques *et al.* 2014). *in vitro* work has further shown that transgenic Dd2 parasites that harboured the 160N allele had significantly higher dihydroartemisinin IC_{50} values compared to those harbouring the wild type S160 allele (Henriques *et al.* 2015). Of interest, the *ap2-mu*-160N mutation was not observed in the MalariaGEN dataset (MalariaGEN Plasmodium falciparum Community Project 2016), particularly in the SE Asian parasites where parasites carry both the CQR *crt* allele and have a multicopy *mdr1* locus. This suggests that this marker may be specific to the African population, even though from the population data there is no evidence of selection. Further evidence showed that transgenic parasites carrying an *ap2-mu*-I592T mutation (not detected in this population), but not those gene-edited to harbour the S160N mutation, had increased ring-stage survival (an *in vitro* proxy for artemisinin resistance) compared to the 3D7 wild type allele after dihydroartemisinin pressure (Henrici, van Schalkwyk and Sutherland 2019). Further work is required to evaluate the potential role of *ap2-mu* mutations on ACT response.

Previous work has shown that artemisinin activity is dependent on haemoglobin uptake and digestion. *falcipain-2a* encodes a cysteine protease that is involved in this pathway and its deletion has been significantly associated with decreased artemisinin sensitivity (Klonis *et al.* 2011). However, while a mutation (S69Stop) that results is artemisinin resistance has been

identified *in vitro* (Ariey *et al.* 2014), this mutation was not identified in our population. However, a significant temporal trend in codon S59F ($\chi 2 = 6.9$, p-value = 0.01) was found and it showed that there was an increase in the wild type S59 allele post-ACT introduction, potentially due to AL pressure. Of note, this mutation is not observed globally (MalariaGEN Plasmodium falciparum Community Project 2016). *falcipain-2a* was the most polymorphic gene in this study and this can be attributed to drug pressure as it is a target of antimalarials (Ponsuwanna *et al.* 2016). Crucially, there were no mutations in codons 185 and 187, two residues within the *falcipain-2a* that are important for binding to haemoglobin (Pasupureddy *et al.* 2019). Recently, Siddiqui et al. (2018) showed that mutant *falcipain-2a* haplotypes together with *k13* artemisinin resistance mutations were associated with elevated survival in the ring-stage survival assay. Some of the mutations identified in that study including N4H, A8I, H10N, Q15H, V51I, S59F, K255R, N257E, T343P, D345G, V393I, A400P and Q414E were identified in this study as well. Though *falcipain-2a* was highly polymorphic, *k13* did not contain artemisinin resistance mutations, thus the role of *falcipain-2a* in modulating artemisinin resistance remains to be understood.

The *ubp-1* 1528D mutation was not identified, contrary to Henriques et al. (2014) and a more recent study conducted in Ghana that found it at a prevalence of 7.4% (Adams *et al.* 2018). Additionally, a recent study identified, among other polymorphisms, a *ubp-1* SNP (R3138H) outside the region genotyped in this study, that was associated with artemisinin resistance on the Thai-Myanmar border (Cerqueira et al. 2017). *ubp-1* was first shown to modulate artemisinin resistance in *P. chabaudi* (Hunt *et al.* 2007) and later in *P. falciparum* (Borrmann *et al.* 2013; Henriques *et al.* 2014). Notably, the frequency of mutant 1528D allele in *P. falciparum* increased after ACT treatment in Western Kenya, although this mutation was not associated with recrudescence (Henriques *et al.* 2014). A recent study demonstrated that death of malaria parasites following artemisinin treatment is achieved through the damage of proteins as well as inhibiting the proteasome (Bridgford *et al.* 2018). Given that *ubp-1* may play a crucial role in ubiquitination and degradation following protein damage, mutations in the genes involved in this pathway may help to enhance artemisinin resistance as

ubiquitination and degradation reduce the accumulation of damaged proteins. Evidently, this calls for further work to understand the role of *ubp-1* as a marker of ACT resistance.

Notably, while there were virtually no significant temporal frequencies in genes other than *crt*, *mdr1* and *dhps*, a distinct pattern was detected from 2005/06 through to 2012/13 and 2015/16. This pattern involved 2-fold increases or decreases in the frequencies of alleles and haplotypes from 2005/06 to 2012/13 and a shift back to the original or similar frequency in 2015/16 and they include: *ap2-mu*-S160N (9.1 - 24 - 16.5%) and E163E (8.51 - 16.36 - 4.55%), *falcipain-2a* N4H (19.2 - 9.2 - 16.5%), H10N (17.6 - 7.1 - 14.6) and E11E (18.9 - 7.1 - 15.3%), *ubp-1* KNE-repeat at codon 1514 (5.5 - 11.6 - 5.41%) and N1518N (5.6 - 11.7 - 5.5%), the second most dominant *k13* haplotype (PK[N6]MAISTLQ, 9.6 - 19.6 - 8.8%) and finally *mdr1*-NFD haplotype with a pattern that goes onto the 2017/18 time point (12.7 - 33.3 - 55 - 41.4 %). This could probably have resulted from the sudden change of drug-policy from SP to ACTs and hence the selection of random SNPs across the genome as the parasite population was adapting to changes in drug-pressure. These SNPs may have been eliminated over time due to a fitness cost.

The major limitation of this study is that factors such as complexity of infection and its role in the evolution of drug-resistance was not studied. The availability of longitudinal samples prompted this study into carrying out a surveillance of anti-malarial drug-resistance markers during periods where first line antimalarials were switched, hence other questions were not explored.

4.5 Conclusion

Following the introduction of ACTs in 2004, there has been a rapid increase in the CQ sensitive population to near fixation and this reignites the debate on the use of CQ for malaria treatment, such as in combination therapy. On the other hand, there is still a need for careful monitoring of the *dhps* A581G locus since SP has proved useful in IPTp, significantly reducing morbidity in pregnant women (van Eijk *et al.* 2019). The decline in the novel marker (*nfs*) which potentially confers resistance to LM, contrary to the observations made in The Gambia, calls for additional studies to determine its role as a potential drug target. The artemisinin

resistance-conferring SE Asian mutations in *k13*, such as C580Y, have not been identified in Kilifi and many of the SNPs occurred in the N-terminal region of the gene with no evidence of drug selection. Consequently, due to lack of the validated molecular markers of artemisinin and lumefantrine resistance, there appears to be no threat to ACT efficacy from resistance in the population, however, continued surveillance remains a requirement.

Chapter 5 : General Discussion

The renewed efforts to eliminate malaria mean that more research is required to inform malaria interventions. Among the impediments to malaria elimination are asymptomatic infections and antimalarial drug-resistance. Asymptomatic infections have been associated with adverse health effects and particularly, provide a constant source of parasites that fuel malaria transmission. This is compounded by the fact that individuals with asymptomatic infections do not present with symptoms of malaria and hence do not seek treatment. In turn, such individuals are often not targeted by malaria interventions such as the use of antimalarial drugs and, hence, remain chronic for long periods. Recent studies have therefore emphasized the need to include asymptomatic infections in malaria interventions and additional research is needed to inform this.

On the other hand, by rendering drugs ineffective, antimalarial drug-resistance makes it difficult to clear malaria infections. Worryingly, drug resistance has constantly emerged against widely used antimalarial drugs and has, in the past, been associated with a significant increase in clinical cases of malaria and mortality. Children under the age of 5 years bear the greatest burden of malaria and antimalarial drug resistance is a problem that does not seem to go away, however, it can be mitigated with informed decisions. Such decisions involve efforts such as carrying out the surveillance of drug resistance using molecular markers. This method has gained prominence given that it is less resource-intensive compared to other methods such as TES and *in vitro* culture of parasites. Additionally, it can include hundreds of samples with ease because dried blood spots can be used and these are easy to collect, ship and store. Thus, the analysis presented in this thesis focussed on characterising *P. falciparum* asymptomatic infections as well as carrying out the surveillance of drug resistance markers in Kilifi, Kenya.

In characterising *P. falciparum* asymptomatic infections, the first objective in chapter 2 focussed on evaluating whether age and malaria transmission intensity impact on asymptomatic infections, hence, predicting the risk of developing febrile malaria. This work was informed by preceding studies that showed divergent outcomes. In some studies, asymptomatic infections were associated with an increased risk of febrile malaria while in others, the inverse was observed. An inspection of these earlier studies revealed that they were conducted in diverse geographical settings of varying malaria transmission intensities

and with participants of varying age groups. Living in regions of high transmission intensity and increasing age have been associated with increasing NAI, which reduces the number of clinical episodes that an individual presents with. For this reason, it would be justified to hypothesize that both transmission intensity and age might have modified the risk of developing febrile malaria in individuals with asymptomatic infections. However, this assumption needed confirmation. To achieve this, malaria monitoring data spanning 19 years was obtained from three cohorts in Kilifi of varying malaria transmission intensities (Ngerenya – low, Junju – moderate to high and Chonyi - high) and with participants ranging between one month to 15 years. In this cohorts, children were recruited from birth until they reached 15 years of age and underwent both active and passive surveillance to detect asymptomatic and febrile infections. Accordingly, over 11,000 person-years of follow up were used to evaluate whether asymptomatic infections predict subsequent febrile malaria infections and whether age and transmission intensity modified this risk.

In the low transmission setting, it was revealed that asymptomatic infections were associated with an increased risk of febrile malaria across all ages, implying that asymptomatic infections are detrimental in this setting, as such individuals are expected to have lower NAI. In the regions of moderate to high transmission, asymptomatic infections were associated with a reduced risk of febrile malaria episodes in older children, however, there was no association in younger children. This implied that asymptomatic infections are protective from malaria episodes and this pointed to a role of higher levels of NAI in such individuals to predict protection. However, regarding the role of premunition or acquired immunity in protecting asymptomatic individuals against illness, this study was not designed to elucidate this. Premunition in asymptomatic individuals would involve a host response that protects against illness due to an ongoing infection, without eliminating the infection. On the other hand, acquired immunity would involve protection against illness due to past infections, still without eliminating the infection. The impact of this is that clearing asymptomatic infections would increase the risk of febrile episodes in the case of premunition but not acquired immunity. Either way, a better understanding of these mechanisms would be instrumental in the design of a malaria vaccine.

Given that this analysis used data from varying transmission intensities and with participants of varying ages, these findings provided a unifying explanation on the role of asymptomatic infections on the risk of developing febrile malaria. Not only do asymptomatic infections predict subsequent febrile malaria infections, but it appears this risk is modified by both transmission intensity and age. As previously mentioned, asymptomatic infections have serious adverse health effects and fuel malaria transmission. Also, while it appears that asymptomatic infections can protect one from malaria in regions of high transmission intensity, they increase one's risk of malaria in regions of low transmission intensity. Finally, there is ample evidence that curing school children of chronic and asymptomatic infections with targeted MDA improves school attendance, cognition and educational attainment (Clarke *et al.* 2017). Therefore, it is crucial that malaria intervention efforts also target asymptomatic infections.

Targeting asymptomatic infections is presented with two main challenges. Firstly, infected individuals do not experience symptoms, hence do not seek treatment and this makes these infections both undetected and likely to become chronic. Secondly, some asymptomatic infections harbour sub-microscopic parasitemia that is below the detection levels of microscopy, the gold standard for malaria diagnosis. Accordingly, this makes the estimation of the burden of asymptomatic infections a major challenge. MDA interventions avoid the detection problem by treating everyone, and some studies have found it to be effective in clearing these infections. However, for long-term impact, it appears that MDA needs to be sustained for longer periods and this may be a challenge in resource-limited settings such as those in SSA. Still, MDA is an attractive method for tackling asymptomatic infections and more research is needed to better deploy MDA. Additionally, MDA is recommended in areas approaching interruption of transmission where there is good access to treatment, effective implementation of vector control and surveillance, as well as regions with minimal risk of re-introduction of infection. Many regions of SSA are yet to achieve significant reductions in levels of malaria transmission, implying that MDA might not be effective in such regions. Therefore, efforts should be geared towards reducing the levels of transmission before MDA can be explored in treating asymptomatic infections in SSA.

The second objective, addressed in chapter 3, was informed by earlier studies that showed that the risk of subsequent febrile malaria episodes following an asymptomatic infection, was due to the introduction of new clones not detected before. However, the characterisation of the genetic diversity of malaria parasites has for a long time relied on *msp1/msp2/glurp* genotyping. These methods, devised in the 1990s, are limited in their sensitivity and are laborious to conduct. Additionally, they are hampered by technical problems such as the use of gel electrophoresis to discriminate parasite clones, something that can be subjective when the bands being analysed overlap. For this reason, earlier studies may have been limited in their ability to fully characterise asymptomatic infections, thereby missing minority clones. If this was the case, the finding that the transition from an asymptomatic episode was due to the introduction of a new clone, may have been hampered by the lack of detection of the same clone in the preceding asymptomatic episode, but at lower frequencies. Amplicon deep sequencing was therefore used to characterise the genetic diversity of parasites in asymptomatic infections as it is a more sensitive tool to detect minority clones and sample pooling allows genotyping of hundreds of samples. Asymptomatic and febrile samples were obtained from the Junju cohort described in the first objective and the *ama1* gene was used to determine COI to characterise parasite diversity as well to characterise the transition from an asymptomatic to febrile malaria episode.

It was revealed that asymptomatic infections harboured more diverse parasites compared to febrile malaria infections similar to previous studies. Markedly, a study conducted in Tanzania found that individuals with asymptomatic infections had higher NAI based or higher antibody titres to merozoite antigens compared to those with febrile infections (Rono *et al.* 2013). In turn, asymptomatic individuals may be expected to harbour more diverse parasites, a proxy of higher NAI. In the transition from an asymptomatic to febrile malaria episode, the results in Chapter 3 suggest that many of the febrile malaria infections were due to a new clone. This finding resonated with earlier studies, implying that the shift from an asymptomatic to febrile malaria episode is caused by a parasite clone to which the host is yet to encounter and hence a higher probability that the parasite will escape the host's immune response. Immunity to malaria involves the generation of antibodies against the parasite, however, the diversity of

the targets of these antibodies impedes the acquisition of sterile immunity. The P. *falciparum* parasite takes full advantage of this diversity to continuously cause clinical episodes and it has been hypothesized that for one to be fully protected from malaria, they would have to encounter multiple diverse parasites. The findings in chapter 3 add weight to this hypothesis and emphasize the need for further studies that will determine the correlates of immunity to malaria. Consequently, such studies will inform the design of a malaria vaccine, a much-needed intervention that could significantly protect from malaria and greatly reduce the human reservoir of disease, as has been observed with the eradication of diseases such as smallpox, due to vaccines.

Another take home point from chapter 3 was the utility of amplicon deep-sequencing to determine COI from a large pool of samples. Previous methods used to characterise malaria transmission intensity such as the use of spleen rate, EIR, annual parasite incidence and parasite prevalence are met with methodological challenges such as being resource-intensive. COI has been shown to be a proxy for malaria transmission intensity and such data can be obtained from hundreds of samples using amplicon deep sequencing. An assessment of the population level COI can help to detect active transmission foci if an increase in COI is detected in an area and it can also be used to detect the introduction of parasites in an area by characterising parasite relatedness. While amplicon deep sequencing is expensive to set up, the cost of sequencing has been decreasing rapidly and the usefulness of this method can be realised in a few years given that hundreds of samples can be pooled for analysis. Notably, one recent review suggested that the high cost of sequencing can be alleviated by creating genomic Centres of Excellence in Africa, which will receive and process samples from a network of smaller laboratories (Apinjoh *et al.* 2019). Consequently, such consortia could create networks and build capacity in SSA for genomics to track and monitor malaria elimination.

If asymptomatic infections were a threat to malaria elimination, then antimalarial drug resistance presents an even bigger threat. The use of drugs to treat malaria, in addition to vector control, has become a cornerstone in malaria control. In the past, the use of CQ and SP as first-line antimalarial drugs has helped to prevent millions of clinical cases and deaths

due to malaria. Thus, lives have been saved, notably those of children under 5 who bear the greatest burden of malaria. Unfortunately, *P. falciparum* has developed resistance to both CQ and SP as well as second-line drugs such as MQ and AQ and this hampers malaria intervention efforts. In the wake of failing CQ, the world changed to SP as the first-line antimalarial drug and later to ACTs after resistance to SP spread widely. Thus, surveillance of antimalarial drug resistance is employed to detect resistance as well as characterising its spread in space and time. The emergence of resistance to artemisinin, a component of ACTs, in SE Asia means that the world is yet again at risk of losing the usefulness of a major first-line antimalarial drug to resistance. Moreover, there is now widespread resistance not only to artemisinin but resistance to ACT partner drugs in SE Asia. This has resulted in countries such as Cambodia to switch from DHA-PPQ to AS-MQ, due to widespread failure of DHA-PPQ to treat malaria infections. From 2000 to 2015, both vector control and the use ACTs have had a major impact in reducing clinical cases of malaria especially in SSA and resistance to ACTs poses a major threat to fighting malaria in SSA.

The identification of a molecular marker of resistance for artemisinin has made it possible to carry out surveillance of artemisinin resistance. Studies have shown that the geographical spread of mutations that correlate with artemisinin resistance in the propeller domain of *k13* overlap with cases of slow clearing parasites. The malaria world has hence moved to contain resistance in SEA, eradicate it and prevent its spread to other malaria endemic regions. *K13* has proven to be a valid marker for achieving this, however, there is a growing concern that *k13* may not be applicable in tracking artemisinin resistance outside SEA. Recent studies have reported cases of parasites that are resistant to artemisinin but lacking *k13* mutations, while other studies have reported mutations in genes other than *k13*. In chapter 4, a temporal analysis of drug resistance was carried out using *P. falciparum* positive samples collected in Kilifi during the pre-ACT period (at a time when CQ and SP were in use) and post-ACT introduction. The main objective was to determine whether the validated SNPs of artemisinin resistance in *k13* exist and whether there has been any selection of mutations in the *k13* gene post-ACT introduction. Additionally, eleven other genes that have been associated with CQ, SP and artemisinin resistance in earlier studies, were included.

The validated SNPs of artemisinin resistance in *k13* were not detected. Also, a *k13* mutation that is prevalent in SSA was found not to be under selection as it was prevalent in the population in the pre-ACT and post-ACT period. Similarly, for the *ap2-mu* S160T that was selected for in children treated with ACTs in Western Kenya, no changes in allele frequencies were observed in the pre- and post-ACT periods. A *falcipain-2a* mutation that was detected in a lab strain that was cultured under artemisinin pressure, was not detected hence in this population there is no evidence of artemisinin resistance from evaluation of known markers. Multiple studies across Africa have continually found the efficacy of ACTs to be high, despite the looming threat of artemisinin resistance in SEA. This is important because it means for now, Africa can continue to rely on ACTs for the treatment of malaria, but should continue to monitor for the emergence of artemisinin resistance. Several theories have been proposed as to why artemisinin resistance is yet to be detected in Africa including *(i)* high parasite diversity and low linkage disequilibrium which would continually break resistant haplotypes due to parasites of diverse background interbreeding. The high efficacy of ACT partner drugs prevents the emergence of resistant parasites since they effectively mop-up the last remaining parasites preventing the exposure of parasite to sub-optimal doses of artemisinin. *(iii)* There is higher acquired immunity to malaria in Africa compared to SEA which would mean that individuals rarely develop illness and hence no need to take drugs and this would substantially reduce drug-pressure then reduce risk of selecting for artemisinin resistance parasites. In line with this theory of immunity, a recent study showed that naturally acquired malarial immunity was associated with reduced anti-malarial treatment failure in malaria endemic populations (O'Flaherty *et al.* 2017). Consequently, it was hypothesized that naturally acquired immunity may accelerate the clearance of artemisinin-resistant parasites and hence immunity may also play an important role in the emergence and transmission potential of artemisinin-resistant parasites. This, however, presents a challenge in that early signs of the emergence of artemisinin resistance parasites may be missed because resistant parasites may emerge and spread under the radar, hence the need for continued surveillance.

The findings in chapter 4 are in line with multiple other studies conducted in SSA that are yet to report the existence of validated artemisinin resistance markers. However, there have been two reports of artemisinin resistant parasites in Equatorial Guinea and Uganda, but these parasites did not have the *k13* SNPs that have been associated with artemisinin resistance. Still, continued surveillance of artemisinin resistance is called for in SSA especially given that there is no immediate replacement for ACTs. Such efforts should involve laboratory studies that aim to generate parasites that are resistant to two widely used ACTs in SSA (AL and AS+AQ) followed by genome-wide association studies to try and detect SNPs under selection that may be involved in resistance. This may help to understand artemisinin resistance in parasites with a SSA genetic background, given that recent studies have reported that artemisinin resistance required a particular genetic background to emerge in SEA.

In conclusion, malaria elimination offers a path to eradicating one of the deadliest diseases known to man. Many gains in tackling malaria have been realised including eliminating malaria in several countries, the deployment of vector control measures and antimalarial drugs that have been key pillars in malaria control as well as substantial increases in global investment in fighting malaria. Still, significant challenges persist, as many regions continue to experience high malaria transmission, insecticide resistance, drug-resistance and chronic asymptomatic infections. In this thesis, I have presented work on the epidemiology and genetic epidemiology of asymptomatic infections and revealed the extensive diversity of parasites that negatively impacts malaria elimination efforts such as the design of a malaria vaccine. Additionally, no evidence of artemisinin resistance was detected in Kilifi, but the foundation for carrying out continuous genetic epidemiology of antimalarial drug resistance in this and other populations continues to be laid. Going forward, baseline data on antimalarial drug-resistance markers from Kilifi are now available and this can be used as the benchmark for continually monitoring the impact of ACTs on the diversity of these markers. Consequently, such data can aid in the detection of early signs of artemisinin resistance. Regarding asymptomatic infections, like previous studies, the current study has shown that parasite diversity is higher in asymptomatic infections compared to febrile infections. Other studies have found this to be as a result of the broader spectrum of immunological memory

in asymptomatic infections which has also been associated with protection from febrile illness. Future work should look to improve our understanding of the mechanisms of protection as this can be used in the design of a malaria vaccine. Given the opportunity, I would incorporate amplicon deep-sequencing into malaria monitoring studies to characterise parasite diversity as well as carry out surveillance of drug-resistance markers. In characterising parasite diversity, a proxy for malaria transmission, studies can evaluate whether malaria interventions efforts are working since they are expected to reduce parasite diversity. Accordingly, such efforts can evaluate whether gains are being made in tackling malaria as we target to eliminate the disease. Secondly, the current study has used Sanger sequencing in the surveillance of drug-resistance markers, however, amplicon deep-sequencing offers a more sensitive, robust and resourceful alternative. With deep-sequencing, many more samples can be incorporated due to pooling, many markers and samples can be included from across Kenya. With this pipeline, routine surveillance can be conducted since the platform is already available at the KEMRI-Wellcome Trust laboratories as well as other sites across Kenya. Such data will keep Kenya ahead of the impending threat of artemisinin resistance since early signs of the emergence of resistance can be detected and mitigated in time.

I hope that the findings presented in this thesis will play a significant role in informing malaria elimination, notably, the utility of genetic epidemiology of malaria and that future studies will further this work to improve our understanding of parasite origins, gene flow between populations and antimalarial drug resistance.

References

Abdel-Latif MS, Dietz K, Issifou S *et al.* Antibodies to *Plasmodium falciparum* Rifin Proteins are Associated with Rapid Parasite Clearance and Asymptomatic Infections. *Infect Immun* 2003;**71**:6229–33.

Achan J, Adam I, Arinaitwe E *et al.* The Effect of Dosing Regimens on the Antimalarial Efficacy of Dihydroartemisinin-Piperaquine: A Pooled Analysis of Individual Patient Data. Garner P (ed.). *PLoS Med* 2013;**10**:e1001564.

Achieng AO, Muiruri P, Ingasia LA *et al.* Temporal trends in prevalence of *Plasmodium falciparum* molecular markers selected for by artemether–lumefantrine treatment in pre-ACT and post-ACT parasites in western Kenya. *Int J Parasitol Drugs Drug Resist* 2015;**5**:92–9.

Adams T, Ennuson NAA, Quashie NB *et al.* Prevalence of Plasmodium falciparum delayed clearance associated polymorphisms in adaptor protein complex 2 mu subunit (*Pfap2mu*) and ubiquitin specific protease 1 (*Pfubp1*) genes in Ghanaian isolates. *Parasit Vectors* 2018;**11**:175.

Adjah J, Fiadzoe B, Ayanful-Torgby R *et al.* Seasonal variations in *Plasmodium falciparum* genetic diversity and multiplicity of infection in asymptomatic children living in southern Ghana. *BMC Infect Dis* 2018;**18**:432.

Agrawal S, Moser KA, Morton L *et al.* Association of a Novel Mutation in the <i>Plasmodium falciparum<i> Chloroquine Resistance Transporter With Decreased Piperaquine Sensitivity. *J Infect Dis* 2017;**216**:468–76.

Al-Yaman F, Genton B, Reeder JC *et al.* Reduced risk of clinical malaria in children infected with multiple clones of *Plasmodium falciparum* in a highly endemic area: a prospective community study. *Trans R Soc Trop Med Hyg* 1997;**91**:602–5.

Alam MT, de Souza DK, Vinayak S *et al.* Selective Sweeps and Genetic Lineages of *Plasmodium falciparum* Drug-Resistant Alleles in Ghana. *J Infect Dis* 2011;**203**:220–7.

Alekseyev YO, Fazeli R, Yang S *et al.* A Next-Generation Sequencing Primer – How Does It Work and What Can It Do? *Acad Pathol* 2018;**5**:237428951876652.

Alves FP, Gil LHS, Marrelli MT *et al.* Asymptomatic carriers of *Plasmodium spp.* as infection

source for malaria vector mosquitoes in the Brazilian Amazon. *J Med Entomol* 2005;**42**:777–9.

Amambua-Ngwa A, Jeffries D, Amato R *et al.* Consistent signatures of selection from genomic analysis of pairs of temporal and spatial *Plasmodium falciparum* populations from The Gambia. *Sci Rep* 2018;**8**:9687.

Amambua-Ngwa A, Okebe J, Mbye H *et al.* Sustained Ex Vivo Susceptibility of *Plasmodium falciparum* to Artemisinin Derivatives but Increasing Tolerance to Artemisinin Combination Therapy Partner Quinolines in The Gambia. *Antimicrob Agents Chemother* 2017;**61**:e00759-17.

Amaratunga C, Lim P, Suon S *et al.* Dihydroartemisinin-piperaquine resistance in *Plasmodium falciparum* malaria in Cambodia: A multisite prospective cohort study. *Lancet Infect Dis* 2016;**16**:357–65.

Amato R, Lim P, Miotto O *et al.* Genetic markers associated with dihydroartemisinin–piperaquine failure in *Plasmodium falciparum* malaria in Cambodia: a genotype–phenotype association study. *Lancet Infect Dis* 2017;**17**:164–73.

Amin AA, Zurovac D, Kangwana BB *et al.* The challenges of changing national malaria drug policy to artemisinin-based combinations in Kenya. *Malar J* 2007;**6**:72.

Anderson TJC, Su XZ, Bockarie M *et al.* Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. *Parasitology* 1999;**119**:S0031182099004552.

Angrisano F, Tan Y-H, Sturm A *et al.* Malaria parasite colonisation of the mosquito midgut-placing the *Plasmodium* ookinete centre stage. *Int J Parasitol* 2012;**42**:519–27.

Apinjoh TO, Ouattara A, Titanji VPK *et al.* Genetic diversity and drug resistance surveillance of *Plasmodium falciparum* for malaria elimination: is there an ideal tool for resource-limited sub-Saharan Africa? *Malar J* 2019;**18**:217.

Ariey F, Witkowski B, Amaratunga C *et al.* A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* 2014;**505**:50–5.

Ashley EA, Dhorda M, Fairhurst RM *et al.* Spread of Artemisinin Resistance in *Plasmodium falciparum* Malaria. *N Engl J Med* 2014;**371**:411–23.

Assefa SA, Preston MD, Campino S *et al.* estMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinformatics* 2014;**30**:1292–4.

Auparakkitanon S, Chapoomram S, Kuaha K *et al.* Targeting of Hematin by the Antimalarial Pyronaridine. *Antimicrob Agents Chemother* 2006;**50**:2197–200.

Autino B, Noris A, Russo R *et al.* Epidemiology of Malaria In Endemic Areas. *Mediterr J Hematol Infect Dis* 2012;**4**:2012060.

Awolola TS, Oyewole IO, Amajoh CN *et al.* Distribution of the molecular forms of *Anopheles gambiae* and pyrethroid knock down resistance gene in Nigeria. *Acta Trop* 2005;**95**:204–9.

Babiker HA, Abdel-Muhsin AM, Ranford-Cartwright LC *et al.* Characteristics of *Plasmodium falciparum* parasites that survive the lengthy dry season in eastern Sudan where malaria transmission is markedly seasonal. *Am J Trop Med Hyg* 1998;**59**:582–90.

Babiker HA, Pringle SJ, Abdel-Muhsin A *et al.* High-level chloroquine resistance in Sudanese isolates of *Plasmodium falciparum* is associated with mutations in the chloroquine resistance transporter gene Pfcrt and the multidrug resistance Gene Pfmdr1. *J Infect Dis* 2001;**183**:1535–8.

Baird JK, Purnomo, Basri H *et al.* Age-specific prevalence of *Plasmodium falciparum* among six populations with limited histories of exposure to endemic malaria. *Am J Trop Med Hyg* 1993;**49**:707–19.

Barry AE, Schultz L, Buckee CO *et al.* Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, *Plasmodium falciparum*. Rénia L (ed.). *PLoS One* 2009;**4**:e8497.

Basco LK, de Pécoulas PE, Wilson CM *et al.* Point mutations in the dihydrofolate reductase-thymidylate synthase gene and pyrimethamine and cycloguanil resistance in *Plasmodium falciparum*. *Mol Biochem Parasitol* 1995;**69**:135–8.

Beck HP, Felger I, Huber W *et al.* Analysis of Multiple *Plasmodium falciparum* Infections in Tanzanian Children during the Phase III Trial of the Malaria Vaccine SPf66. *J Infect Dis* 1997;**175**:921–6.

de Beer TAP, Louw AI, Joubert F. Elucidation of sulfadoxine resistance with structural models of the bifunctional *Plasmodium falciparum* dihydropterin pyrophosphokinase-dihydropteroate synthase. *Bioorg Med Chem* 2006;**14**:4433–43.

Bei AK, Niang M, Deme AB *et al.* Dramatic Changes in Malaria Population Genetic Complexity in Dielmo and Ndiop, Senegal, Revealed Using Genomic Surveillance. *J Infect Dis* 2018;**217**:622–7.

Beier JC, Killeen GF, Githure JI. Short report: Entomologic inoculation rates and *Plasmodium falciparum* malaria prevalence in Africa. *Am J Trop Med Hyg* 1999;**61**:109–13.

Bejon P, Andrews L, Hunt-Cooke A *et al.* Thick blood film examination for *Plasmodium falciparum* malaria has reduced sensitivity and underestimates parasite density. *Malar J* 2006;**5**:104.

Bejon P, Williams TN, Liljander A *et al.* Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med* 2010;**7**:e1000304.

Bell AS, Roode JC, Sim D *et al.* Within-host competition in genetically diverse malaria infections: parasite virulence and competitive success. *Evolution (N Y)* 2006;**60**:1358–71.

Bendixen M, Msangeni HA, Pedersen BV *et al.* Diversity of *Plasmodium falciparum* populations and complexity of infections in relation to transmission intensity and host age: a study from the Usambara Mountains, Tanzania. *Trans R Soc Trop Med Hyg* 2001;**95**:143–8.

Bereczky S, Liljander A, Rooth I *et al.* Multiclonal asymptomatic *Plasmodium falciparum* infections predict a reduced risk of malaria disease in a Tanzanian population. *Microbes Infect* 2007;**9**:103–10.

Beshir KB, Sutherland CJ, Sawa P *et al.* Residual *Plasmodium falciparum* parasitemia in Kenyan children after artemisinin-combination therapy is associated with increased

transmission to mosquitoes and parasite recurrence. *J Infect Dis* 2013;**208**:2017–24.

Bhatt S, Weiss DJ, Cameron E *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* 2015;**526**:207–11.

Birnbaum J, Scharf S, Schmidt S *et al.* A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites. *Science (80-)* 2020;**367**:51–9.

Borrmann S, Straimer J, Mwai L *et al.* Genome-wide screen identifies new candidate genes associated with artemisinin susceptibility in Plasmodium falciparum in Kenya. *Sci Rep* 2013;**3**:3318.

Van Bortel W, Trung HD, Hoi LX *et al.* Malaria transmission and vector behaviour in a forested malaria focus in central Vietnam and the implications for vector control. *Malar J* 2010;**9**:373.

Bousema T, Drakeley C. Epidemiology and Infectivity of *Plasmodium falciparum* and *Plasmodium vivax* Gametocytes in Relation to Malaria Control and Elimination. *Clin Microbiol Rev* 2011;**24**:377–410.

Bousema T, Okell L, Felger I *et al.* Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat Rev Microbiol* 2014;**12**:833–40.

Boussaroque A, Fall B, Madamet M *et al.* Emergence of Mutations in the K13 Propeller Gene of *Plasmodium falciparum* Isolates from Dakar, Senegal, in 2013-2014. *Antimicrob Agents Chemother* 2016;**60**:624–7.

Branch OH, Takala S, Kariuki S *et al. Plasmodium falciparum* genotypes, low complexity of infection, and resistance to subsequent malaria in participants in the Asembo Bay Cohort Project. *Infect Immun* 2001;**69**:7783–92.

Breman JG. The ears of the hippopotamus: Manifestations, determinants, and estimates of the malaria burden. *American Journal of Tropical Medicine and Hygiene*. Vol 64. 2001, 1–11.

Bridgford JL, Xie SC, Cobbold SA *et al.* Artemisinin kills malaria parasites by damaging proteins and inhibiting the proteasome. *Nat Commun* 2018;**9**:3801.

Buchwald AG, Sixpence A, Chimenya M *et al.* Clinical Implications of Asymptomatic *Plasmodium falciparum* Infections in Malawi. *Clin Infect Dis* 2018;**68**:106–12.

Bzik DJ, Li WB, Horii T *et al.* Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene. *Proc Natl Acad Sci* 1987;**84**:8360–4.

Campino S, Auburn S, Kivinen K *et al.* Population genetic analysis of *Plasmodium falciparum* parasites using a customized illumina goldengate genotyping assay. Fitzgerald JR (ed.). *PLoS One* 2011;**6**:e20251.

Carneiro I, Roca-Feltrer A, Griffin JT *et al.* Age-Patterns of Malaria Vary with Severity, Transmission Intensity and Seasonality in Sub-Saharan Africa: A Systematic Review and Pooled Analysis. Noor AM (ed.). *PLoS One* 2010;**5**:e8988.

Cerqueira GC, Cheeseman IH, Schaffner SF *et al.* Longitudinal genomic surveillance of Plasmodium falciparum malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. *Genome Biol* 2017;**18**:78.

Chan J-A, Howell KB, Reiling L *et al.* Targets of antibodies against Plasmodium falciparum–infected erythrocytes in malaria immunity. *J Clin Invest* 2012;**122**:3227–38.

Chattopadhyay R, Srivastava VK, Das BS *et al.* Epidemiology of malaria transmission and development of natural immunity in a malaria-endemic village, San Dulakudar, in Orissa state, India. *Am J Trop Med Hyg* 2004;**71**:457–65.

Chaumeau V, Fustec B, Nay Hsel S *et al.* Entomological determinants of malaria transmission in Kayin state, Eastern Myanmar: A 24-month longitudinal study in four villages. *Wellcome Open Res* 2019;**3**:109.

Chen I, Clarke SE, Gosling R *et al.* "Asymptomatic" Malaria: A Chronic and Debilitating Infection That Should Be Treated. *PLOS Med* 2016;**13**:e1001942.

Chen J-T, Li J, Zha G-C *et al.* Genetic diversity and allele frequencies of *Plasmodium falciparum msp1* and *msp2* in parasite isolates from Bioko Island, Equatorial Guinea. *Malar J* 2018;**17**:458.

Chenet SM, Okoth SA, Kelley J *et al.* Molecular Profile of Malaria Drug Resistance Markers of *Plasmodium falciparum* in Suriname. *Antimicrob Agents Chemother* 2017;**61**:e02655-16.

Chiodini PL. Malaria diagnostics: now and the future. *Parasitology* 2014;**141**:1873–9.

Churcher TS, Lissenden N, Griffin JT *et al.* The impact of pyrethroid resistance on the efficacy and effectiveness of bednets for malaria control in Africa. *Elife* 2016;**5**, DOI: 10.7554/eLife.16090.

Clarke SE, Rouhani S, Diarra S *et al.* Impact of a malaria intervention package in schools on *Plasmodium* infection, anaemia and cognitive function in schoolchildren in Mali: a pragmatic cluster-randomised trial. *BMJ Glob Heal* 2017;**2**:e000182.

Cogswell FB. The hypnozoite and relapse in primate malaria. *Clin Microbiol Rev* 1992;**5**:26–35.

Coldiron ME, Von Seidlein L, Grais RF. Seasonal malaria chemoprevention: successes and missed opportunities. *Malar J* 2017;**16**:481.

Coleman M, Hemingway J, Gleave KA *et al.* Developing global maps of insecticide resistance risk to improve vector control. *Malar J* 2017;**16**:86.

Coleman RE, Kumpitak C, Ponlawat A *et al.* Infectivity of Asymptomatic *Plasmodium* -Infected Human Populations to Anopheles dirus Mosquitoes in Western Thailand. *J Med Entomol* 2004;**41**:201–8.

Collins WE, Jeffery GM. *Plasmodium ovale*: Parasite and Disease. *Clin Microbiol Rev* 2005;**18**:570–81.

Collins WE, Jeffery GM. *Plasmodium malariae*: Parasite and Disease. *Clin Microbiol Rev* 2007;**20**:579–92.

Combrinck JM, Mabotha TE, Ncokazi KK *et al.* Insights into the Role of Heme in the Mechanism of Action of Antimalarials. *ACS Chem Biol* 2013;**8**:133–7.

Conrad MD, Bigira V, Kapisi J *et al.* Polymorphisms in *k13* and *falcipain-2* associated with artemisinin resistance are not prevalent in *Plasmodium falciparum* isolated from

Ugandan children. Tsuboi T (ed.). *PLoS One* 2014;**9**:e105690.

Contamin H, Fandeur T, Rogier C *et al.* Different genetic characteristics of *Plasmodium falciparum* isolates collected during successive clinical malaria episodes in Senegalese children. *Am J Trop Med Hyg* 1996;**54**:632–43.

Corbel V, N'Guessan R. Distribution, Mechanisms, Impact and Management of Insecticide Resistance in Malaria Vectors: A Pragmatic Review. *Anopheles Mosquitoes - New Insights into Malaria Vectors*. InTech, 2013.

Cottrell G, Moussiliou A, Luty AJF *et al.* Submicroscopic Plasmodium falciparum Infections Are Associated With Maternal Anemia, Premature Births, and Low Birth Weight. *Clin Infect Dis* 2015;**60**:1481–8.

Coura JR, Suárez-Mutis M, Ladeia-Andrade S. A new challenge for malaria control in Brazil: asymptomatic *Plasmodium* infection - a review. *Mem Inst Oswaldo Cruz* 2006;**101**:229–37.

Cowman AF. A P-glycoprotein homologue of *Plasmodium falciparum* is localized on the digestive vacuole. *J Cell Biol* 1991;**113**:1033–42.

Cowman AF, Galatis D, Thompson JK. Selection for mefloquine resistance in *Plasmodium falciparum* is linked to amplification of the *Pfmdr1* gene and cross-resistance to halofantrine and quinine. *Proc Natl Acad Sci* 1994;**91**:1143–7.

Cox MJ, Kum DE, Tavul L *et al.* Dynamics of malaria parasitaemia associated with febrile illness in children from a rural area of Madang, Papua New Guinea. *Trans R Soc Trop Med Hyg* 1994;**88**:191–7.

Cravo P, Napolitano H, Culleton R. How genomics is contributing to the fight against artemisinin-resistant malaria parasites. *Acta Trop* 2015;**148**:1–7.

Culleton RL, Mita T, Ndounga M *et al.* Failure to detect *Plasmodium vivax* in West and Central Africa by PCR species typing. *Malar J* 2008;**7**:174.

Daniels R, Volkman SK, Milner DA *et al.* A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. *Malar J* 2008;**7**:223.

Daniels RF, Schaffner SF, Wenger EA *et al.* Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proc Natl Acad Sci* 2015;**112**:7067–72.

Daubersies P, Sallenave-Sales S, Magne S *et al.* Rapid turnover of *Plasmodium falciparum* populations in asymptomatic individuals living in a high transmission area. *Am J Trop Med Hyg* 1996;**54**:18–26.

Demas AR, Sharma AI, Wong W *et al.* Mutations in *Plasmodium falciparum* actin-binding protein coronin confer reduced artemisinin susceptibility. *Proc Natl Acad Sci* 2018;**115**:12799–804.

Dhingra SK, Redhi D, Combrinck JM *et al.* A Variant PfCRT Isoform Can Contribute to *Plasmodium falciparum* Resistance to the First-Line Partner Drug Piperaquine. Wellems TE (ed.). *MBio* 2017;**8**:1–19.

Dieye T, Ahouidi AD, Nwakanma D *et al.* West Africa International Centers of Excellence for Malaria Research: Drug Resistance Patterns to Artemether–Lumefantrine in Senegal, Mali, and The Gambia. *Am J Trop Med Hyg* 2016;**95**:1054–60.

Djimdé A, Doumbo OK, Cortese JF *et al.* A Molecular Marker for Chloroquine-Resistant Falciparum Malaria. *N Engl J Med* 2001;**344**:257–63.

Dodoo D, Staalsoe T, Giha H *et al.* Antibodies to Variant Antigens on the Surfaces of Infected Erythrocytes Are Associated with Protection from Malaria in Ghanaian Children. *Infect Immun* 2001;**69**:3713–8.

Dondorp AM, Nosten F, Yi P *et al.* Artemisinin Resistance in *Plasmodium falciparum* Malaria. *N Engl J Med* 2009;**361**:455–67.

Dondorp AM, Yeung S, White L *et al.* Artemisinin resistance: current status and scenarios for containment. *Nat Rev Microbiol* 2010;**8**:272–80.

Doolan DL, Dobaño C, Baird JK. Acquired Immunity to Malaria. *Clin Microbiol Rev* 2009;**22**:13–36.

Dowle M, Srinivasan A. data.table: Extension of `data.frame`. 2018.

Duraisingh MT, Drakeley CJ, Muller O *et al.* Evidence for selection for the tyrosine-86 allele of the pfmdr1 gene of *Plasmodium falciparum* by chloroquine and amodiaquine. *Parasitology* 1997;**114**:205–11.

Durand R, Ariey F, Cojean S *et al.* Analysis of circulating populations of Plasmodium falciparum in mild and severe malaria in two different epidemiological patterns in Madagascar. *Trop Med Int Heal* 2008;**13**:1392–9.

Earland D, Buchwald AG, Sixpence A *et al.* Impact of Multiplicity of *Plasmodium falciparum* Infection on Clinical Disease in Malawi. *Am J Trop Med Hyg* 2019;**101**:412–5.

van Eijk AM, Larsen DA, Kayentao K *et al.* Effect of Plasmodium falciparum sulfadoxine-pyrimethamine resistance on the effectiveness of intermittent preventive therapy for malaria in pregnancy in Africa: a systematic review and meta-analysis. *Lancet Infect Dis* 2019;**3099**:1–11.

Eisele TP, Bennett A, Silumbe K *et al.* Short-term Impact of Mass Drug Administration With Dihydroartemisinin Plus Piperaquine on Malaria in Southern Province Zambia: A Cluster-Randomized Controlled Trial. *J Infect Dis* 2016;**214**:1831–9.

Eyase FL, Akala HM, Ingasia L *et al.* The Role of *Pfmdr1* and *Pfcrt* in Changing Chloroquine, Amodiaquine, Mefloquine and Lumefantrine Susceptibility in Western-Kenya P. falciparum Samples during 2008–2011. Templeton TJ (ed.). *PLoS One* 2013;**8**:e64299.

Färnert A, Rooth I, Svensson Å *et al.* Complexity of *Plasmodium falciparum* Infections Is Consistent over Time and Protects against Clinical Disease in Tanzanian Children. *J Infect Dis* 1999;**179**:989–95.

Färnert A, Williams TN, Mwangi TW *et al.* Transmission-Dependent Tolerance to Multiclonal Plasmodium falciparum Infection. *J Infect Dis* 2009;**200**:1166–75.

Farooq U, Mahajan RC. Drug resistance in malaria. *J Vector Borne Dis* 2004;**41**:45–53.

Felger I, Smith T, Edoh D *et al.* Multiple Plasmodium falciparum infections in Tanzanian infants. *Trans R Soc Trop Med Hyg* 1999;**93**:29–34.

Fidock DA, Nomura T, Talley AK *et al.* Mutations in the P. falciparum digestive vacuole

transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol Cell* 2000;**6**:861–71.

Fitch CD. Ferriprotoporphyrin IX, phospholipids, and the antimalarial actions of quinoline drugs. *Life Sci* 2004;**74**:1957–72.

Fogh S, Jepsen S, Effersøe P. Chloroquine-resistant *Plasmodium falciparum* malaria in Kenya. *Trans R Soc Trop Med Hyg* 1979;**73**:228–9.

Fola AA, Harrison GLA, Hazairin MH *et al.* Higher complexity of infection and genetic diversity of *Plasmodium vivax* than *Plasmodium falciparum* across all malaria transmission zones of Papua New Guinea. *Am J Trop Med Hyg* 2017;**96**:630–41.

Fola AA, Nate E, Abby Harrison GL *et al.* Nationwide genetic surveillance of *Plasmodium vivax* in Papua New Guinea reveals heterogeneous transmission dynamics and routes of migration amongst subdivided populations. *Infect Genet Evol* 2018;**58**:83–95.

Foote SJ, Galatis D, Cowman AF. Amino acids in the dihydrofolate reductase-thymidylate synthase gene of *Plasmodium falciparum* involved in cycloguanil resistance differ from those involved in pyrimethamine resistance. *Proc Natl Acad Sci* 1990;**87**:3014–7.

Foote SJ, Kyle DE, Martin RK *et al.* Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in *Plasmodium falciparum*. *Nature* 1990;**345**:255–8.

Foote SJ, Thompson JK, Cowman AF *et al.* Amplification of the multidrug resistance gene in some chloroquine-resistant isolates of *P. falciparum*. *Cell* 1989;**57**:921–30.

Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the sequenom massARRAY iPLEX Platform. *Curr Protoc Hum Genet* 2009:1–18.

Gardner MJ, Hall N, Fung E *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002;**419**:498–511.

Gething PW, Casey DC, Weiss DJ *et al.* Mapping *Plasmodium falciparum* Mortality in Africa between 1990 and 2015. *N Engl J Med* 2016;**375**:2435–45.

Ghani AC, Sutherland CJ, Riley EM *et al.* Loss of Population Levels of Immunity to Malaria as a

Result of Exposure-Reducing Interventions: Consequences for Interpretation of Disease Trends. Riley S (ed.). *PLoS One* 2009;**4**:e4383.

Graves PM, Choi L, Gelband H *et al.* Primaquine or other 8-aminoquinolines for reducing *Plasmodium falciparum* transmission. *Cochrane Database Syst Rev* 2018, DOI: 10.1002/14651858.CD008152.pub5.

Greenhouse B, Dokomajilar C, Hubbard A *et al.* Impact of Transmission Intensity on the Accuracy of Genotyping To Distinguish Recrudescence from New Infection in Antimalarial Clinical Trials. *Antimicrob Agents Chemother* 2007;**51**:3096–103.

Griffin JT, Ferguson NM, Ghani AC. Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. *Nat Commun* 2014;**5**:3136.

Guerra CA, Gikandi PW, Tatem AJ *et al.* The Limits and Intensity of *Plasmodium falciparum* Transmission: Implications for Malaria Control and Elimination Worldwide. Rogerson SJ (ed.). *PLoS Med* 2008;**5**:e38.

Haas BJ, Gevers D, Earl AM *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* 2011;**21**:494–504.

Hamilton WL, Amato R, van der Pluijm RW *et al.* Evolution and expansion of multidrug-resistant malaria in southeast Asia: a genomic epidemiology study. *Lancet Infect Dis* 2019;**3099**:10–5.

Harris I, Sharrock WW, Bain LM *et al.* A large proportion of asymptomatic *Plasmodium* infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. *Malar J* 2010;**9**:254.

Haruki K, Bray PG, Ward SA *et al.* Chloroquine resistance of *Plasmodium falciparum*: further evidence for a lack of association with mutations of the *Pf*mdr1 gene. *Trans R Soc Trop Med Hyg* 1994;**88**:694.

Hathaway NJ, Parobek CM, Juliano JJ *et al.* SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic Acids Res* 2018;**46**:e21–e21.

Henning L, Schellenberg D, Smith T *et al.* A prospective study of *Plasmodium falciparum* multiplicity of infection and morbidity in Tanzanian children. *Trans R Soc Trop Med Hyg* 2004;**98**:687–94.

Henrici RC, van Schalkwyk DA, Sutherland CJ. Modification of *Pfap2μ* and *Pfubp1* Markedly Reduces Ring-Stage Susceptibility of *Plasmodium falciparum* to Artemisinin In Vitro. *Antimicrob Agents Chemother* 2019;**64**, DOI: 10.1128/AAC.01542-19.

Henriques G, Hallett RL, Beshir KB *et al.* Directional selection at the Pfmdr1, Pfcrt, Pfubp1, and Pfap2mu loci of Plasmodium falciparum in Kenyan children treated with ACT. *J Infect Dis* 2014;**210**:2001–8.

Henriques G, van Schalkwyk DA, Burrow R *et al.* The Mu Subunit of *Plasmodium falciparum* Clathrin-Associated Adaptor Protein 2 Modulates In Vitro Parasite Response to Artemisinin and Quinine. *Antimicrob Agents Chemother* 2015;**59**:2540–7.

Holmgren G, Gil JP, Ferreira PM *et al.* Amodiaquine resistant *Plasmodium falciparum* malaria in vivo is associated with selection of pfcrt 76T and pfmdr1 86Y. *Infect Genet Evol* 2006;**6**:309–14.

Hui L, DelMonte T, Ranade K. Genotyping Using the TaqMan Assay. *Current Protocols in Human Genetics*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008.

Humphreys GS, Merinopoulos I, Ahmed J *et al.* Amodiaquine and Artemether-Lumefantrine Select Distinct Alleles of the Plasmodium falciparum mdr1 Gene in Tanzanian Children Treated for Uncomplicated Malaria. *Antimicrob Agents Chemother* 2007;**51**:991–7.

Hunt P, Afonso A, Creasey A *et al.* Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. *Mol Microbiol* 2007;**65**:27–40.

Ikeda M, Kaneko M, Tachibana S-I *et al.* Artemisinin-Resistant Plasmodium falciparum with High Survival Rates, Uganda, 2014–2016. *Emerg Infect Dis* 2018;**24**:718–26.

Imwong M, Suwannasin K, Kunasol C *et al.* The spread of artemisinin-resistant Plasmodium falciparum in the Greater Mekong subregion: a molecular epidemiology observational

study. *Lancet Infect Dis* 2017;**17**:491–7.

Ismail HM, Barton V, Phanchana M *et al.* Artemisinin activity-based probes identify multiple molecular targets within the asexual stage of the malaria parasites *Plasmodium falciparum* 3D7. *Proc Natl Acad Sci* 2016;**113**:2080–5.

Juliano JJ, Porter K, Mwapasa V *et al.* Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proc Natl Acad Sci* 2010;**107**:20138–43.

Kadri S, Zhen CJ, Wurst MN *et al.* Amplicon Indel Hunter Is a Novel Bioinformatics Tool to Detect Large Somatic Insertion/Deletion Mutations in Amplicon-Based Next-Generation Sequencing Data. *J Mol Diagnostics* 2015;**17**:635–43.

Kafsack BFC, Rovira-Graells N, Clark TG *et al.* A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* 2014;**507**:248–52.

Kakuru A, Jagannathan P, Muhindo MK *et al.* Dihydroartemisinin–Piperaquine for the Prevention of Malaria in Pregnancy. *N Engl J Med* 2016;**374**:928–39.

Kamau E, Campino S, Amenga-Etego L *et al.* K13-Propeller Polymorphisms in *Plasmodium falciparum* Parasites From Sub-Saharan Africa. *J Infect Dis* 2015;**211**:1352–5.

Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots. 2020.

Kassambara A, Kosinski M. survminer: Drawing Survival Curves using “ggplot2.” 2018.

Kateera F, Nsobya SL, Tukwasibwe S *et al.* Malaria case clinical profiles and *Plasmodium falciparum* parasite genetic diversity: a cross sectional survey at two sites of different malaria transmission intensities in Rwanda. *Malar J* 2016;**15**:237.

Kaur K, Jain M, Reddy RP *et al.* Quinolines and structurally related heterocycles as antimalarials. *Eur J Med Chem* 2010;**45**:3245–64.

Kelly-Hope LA, McKenzie FE. The multiplicity of malaria transmission: a review of entomological inoculation rate measurements and methods across sub-Saharan Africa. *Malar J* 2009;**8**:19.

Kenea O, Balkew M, Tekie H *et al.* Human-biting activities of *Anopheles species* in south-central Ethiopia. *Parasit Vectors* 2016;**9**:527.

Kleinschmidt I, Bradley J, Knox TB *et al.* Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: a WHO-coordinated, prospective, international, observational cohort study. *Lancet Infect Dis* 2018;**18**:640–9.

Klonis N, Crespo-Ortiz MP, Bottova I *et al.* Artemisinin activity against Plasmodium falciparum requires hemoglobin uptake and digestion. *Proc Natl Acad Sci* 2011;**108**:11405–10.

Korenromp EL, Williams BG, Gouws E *et al.* Measurement of trends in childhood malaria mortality in Africa: An assessment of progress toward targets based on verbal autopsy. *Lancet Infect Dis* 2003;**3**:349–58.

Krieger H, Durlacher RR, Menezes MJ *et al.* High prevalence of asymptomatic *Plasmodium vivax* and *Plasmodium falciparum* infections in native Amazonian populations. *Am J Trop Med Hyg* 2002;**66**:641–8.

Kublin JG, Cortese JF, Njunju EM *et al.* Reemergence of chloroquine-sensitive Plasmodium falciparum malaria after cessation of chloroquine use in Malawi. *J Infect Dis* 2003;**187**:1870–5.

Kun JFJ, Missinou M a., Lell B *et al.* New emerging *Plasmodium falciparum* genotypes in children during the transition phase from asymptomatic parasitemia to malaria. *Am J Trop Med Hyg* 2002;**66**:653–8.

Kwansomboon N, Chaumeau V, Kittiphanakun P *et al.* Vector bionomics and malaria transmission along the Thailand-Myanmar border: a baseline entomological survey. *J Vector Ecol* 2017;**42**:84–93.

Kwiatkowski D. Cytokines and anti-disease immunity to malaria. *Res Immunol* 1991;**142**:707–12.

Landier J, Parker DM, Thu AM *et al.* The role of early detection and treatment in malaria elimination. *Malar J* 2016;**15**:363.

Larsen DA, Bennett A, Silumbe K *et al.* Population-wide malaria testing and treatment with

rapid diagnostic tests and artemether-lumefantrine in Southern Zambia: A community Randomized Step-Wedge Control Trial Design. *Am J Trop Med Hyg* 2015;**92**:913–21.

Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 2014;**30**:3276–8.

Leang R, Mairet-Khedim M, Chea H *et al.* Efficacy and Safety of Pyronaridine-Artesunate plus Single-Dose Primaquine for Treatment of Uncomplicated *Plasmodium falciparum* Malaria in Eastern Cambodia. *Antimicrob Agents Chemother* 2019;**63**, DOI: 10.1128/AAC.02242-18.

Leoratti FMS, Durlacher RR, Lacerda MVG *et al.* Pattern of humoral immune response to *Plasmodium falciparum* blood stages in individuals presenting different clinical expressions of malaria. *Malar J* 2008;**7**:186.

Lerch A, Koepfli C, Hofmann NE *et al.* Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. *BMC Genomics* 2017;**18**:864.

Lewontin RC. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 1964;**49**:49–67.

Liljander A, Bejon P, Mwacharo J *et al.* Clearance of asymptomatic P. falciparum infections interacts with the number of clones to predict the risk of subsequent malaria in Kenyan children. *PLoS One* 2011;**6**:1–8.

Liljander A, Chandramohan D, Kweku M *et al.* Influences of Intermittent Preventive Treatment and Persistent Multiclonal Plasmodium falciparum Infections on Clinical Malaria Risk. *PLoS One* 2010;**5**:e13649.

Lim SS, Fullman N, Stokes A *et al.* Net Benefits: A Multicountry Analysis of Observational Data Examining Associations between Insecticide-Treated Mosquito Nets and Health Outcomes. Noor AM (ed.). *PLoS Med* 2011;**8**:e1001091.

Lin JT, Hathaway NJ, Saunders DL *et al.* Using Amplicon Deep Sequencing to Detect Genetic Signatures of Plasmodium vivax Relapse. *J Infect Dis* 2015;**212**:999–1008.

Lindblade K a, Steinhardt L, Samuels A *et al.* The silent threat: asymptomatic parasitemia and malaria transmission. *Expert Rev Anti Infect Ther* 2013;**11**:623–39.

Lu F, Culleton R, Zhang M *et al.* Emergence of Indigenous Artemisinin-Resistant *Plasmodium falciparum* in Africa. *N Engl J Med* 2017;**376**:991–3.

Lubis IND, Wijaya H, Lubis M *et al.* Contribution of *Plasmodium knowlesi* to Multispecies Human Malaria Infections in North Sumatera, Indonesia. *J Infect Dis* 2017;**215**:1148–55.

Madamet M, Briolant S, Amalvict R *et al.* The *Plasmodium falciparum* chloroquine resistance transporter is associated with the *ex vivo* P. falciparum African parasite response to pyronaridine. *Parasit Vectors* 2016;**9**:77.

Magesa SM, Mdira KY, Babiker HA *et al.* Diversity of *Plasmodium falciparum* clones infecting children living in a holoendemic area in north-eastern Tanzania. *Acta Trop* 2002;**84**:83–92.

Mahajan B, Zheng H, Pham PT *et al.* Polymerase chain reaction-based tests for pan-species and species-specific detection of human *Plasmodium* parasites. *Transfusion* 2012;**52**:1949–56.

Maitland K, Olupot-Olupot P, Kiguli S *et al.* Transfusion Volume for Children with Severe Anemia in Africa. *N Engl J Med* 2019;**381**:420–31.

Maketa V, Mavoko H, da Luz R *et al.* The relationship between Plasmodium infection, anaemia and nutritional status in asymptomatic children aged under five years living in stable transmission zones in Kinshasa, Democratic Republic of Congo. *Malar J* 2015;**14**:83.

MalariaGEN Plasmodium falciparum Community Project. Genomic epidemiology of artemisinin resistant malaria. *Elife* 2016;**5**:1–29.

Males S, Gaye O, Garcia A. Long-Term Asymptomatic Carriage of Plasmodium falciparum Protects from Malaria Attacks: a Prospective Study among Senegalese Children. *Clin Infect Dis* 2008;**46**:516–22.

Mang'era CM, Mbai FN, Omedo IA *et al.* Changes in genotypes of Plasmodium falciparum human malaria parasite following withdrawal of chloroquine in Tiwi, Kenya. *Acta Trop*

2012;**123**:202–7.

Manske M, Miotto O, Campino S *et al.* Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* 2012;**487**:375–9.

Marsh K, Kinyanjui S. Immune effector mechanisms in malaria. *Parasite Immunol* 2006;**28**:51–60.

Martin RE, Shafik SH, Richards SN. Mechanisms of resistance to the partner drugs of artemisinin in the malaria parasite. *Curr Opin Pharmacol* 2018;**42**:71–80.

Mason Dentinger R. Patterns of Infection and Patterns of Evolution: How a Malaria Parasite Brought "Monkeys and Man" Closer Together in the 1960s. *J Hist Biol* 2016;**49**:359–95.

de Mast Q, Brouwers J, Syafruddin D *et al.* Is asymptomatic malaria really asymptomatic? Hematological, vascular and inflammatory effects of asymptomatic malaria parasitemia. *J Infect* 2015;**71**:587–96.

Matangila JR, Lufuluabo J, Ibalanky AL *et al.* Asymptomatic Plasmodium falciparum infection is associated with anaemia in pregnancy and can be more cost-effectively detected by rapid diagnostic test than by microscopy in Kinshasa, Democratic Republic of the Congo. *Malar J* 2014;**13**:132.

May J, Falusi AG, Mockenhaupt FP *et al.* Impact of subpatent multi-species and multi-clonal *Plasmodial* infections on anaemia in children from Nigeria. *Trans R Soc Trop Med Hyg* 2000;**94**:399–403.

Medicines for Malaria Venture. Pyramax® (pyronaridine-artesunate).

Ménard D, Khim N, Beghain J *et al.* A Worldwide Map of Plasmodium falciparum K13-Propeller Polymorphisms. *N Engl J Med* 2016;**374**:2453–64.

Michalakis Y, Renaud F. Evolution in vector control. *Nature* 2009;**462**:298–300.

Mideo N, Bailey JA, Hathaway NJ *et al.* A deep sequencing tool for partitioning clearance rates following antimalarial treatment in polyclonal infections. *Evol Med Public Heal* 2016;**2016**:21–36.

Miotto O, Almagro-Garcia J, Manske M *et al.* Multiple populations of artemisinin-resistant <i>Plasmodium falciparum<i/> in Cambodia. *Nat Genet* 2013;**45**:648–55.

Miotto O, Amato R, Ashley EA *et al.* Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet* 2015;**47**:226–34.

Missinou MA, Kun JF, Lell B *et al.* Change in *Plasmodium falciparum* genotype during successive malaria episodes in Gabonese children. *Parasitol Res* 2001;**87**:1020–3.

Mita T, Venkatesan M, Ohashi J *et al.* Limited geographical origin and global spread of sulfadoxine-resistant dhps alleles in plasmodium falciparum populations. *J Infect Dis* 2011;**204**:1980–8.

Mogeni P, Williams TN, Fegan G *et al.* Age, Spatial, and Temporal Variations in Hospital Admissions with Malaria in Kilifi County, Kenya: A 25-Year Longitudinal Observational Study. Grais RF (ed.). *PLOS Med* 2016;**13**:e1002047.

Mohammed H, Mindaye T, Belayneh M *et al.* Genetic diversity of *Plasmodium falciparum* isolates based on *MSP-1* and *MSP-2* genes from Kolla-Shele area, Arbaminch Zuria District, southwest Ethiopia. *Malar J* 2015;**14**:73.

Mohapatra PK, Prakash A, Bhattacharyya DR *et al.* Detection & molecular confirmation of a focus of *Plasmodium malariae* in Arunachal Pradesh, India. *Indian J Med Res* 2008;**128**:52–6.

Molineaux L, Gramiccia G. *The Garki Project : Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa*. World Health Organisation. https://apps.who.int/iris/handle/10665/40316, 1980.

Mota MM, Pradel G, Vanderberg JP *et al.* Migration of Plasmodium sporozoites through cells before infection. *Science (80-)* 2001;**291**:141–4.

Mueller I, Schoepflin S, Smith T a. *et al.* Force of infection is key to understanding the epidemiology of Plasmodium falciparum malaria in Papua New Guinean children. *Proc Natl Acad Sci* 2012;**109**:10030–5.

Mueller I, Tulloch J, Marfurt J *et al.* Malaria control in Papua New Guinea results in complex

epidemiological changes. *P N G Med J* 2005;**48**:151–7.

Mugittu K, Adjuik M, Snounou G *et al.* Molecular genotyping to distinguish between recrudescents and new infections in treatment trials of *Plasmodium falciparum* malaria conducted in Sub-Saharan Africa: adjustment of parasitological outcomes and assessment of genotyping effectiveness. *Trop Med Int Heal* 2006;**11**:1350–9.

Mukherjee A, Bopp S, Magistrado P *et al.* Artemisinin resistance without pfkelch13 mutations in Plasmodium falciparum isolates from Cambodia. *Malar J* 2017;**16**:195.

Musuva A, Ejersa W, Kiptui R *et al.* The malaria testing and treatment landscape in Kenya: results from a nationally representative survey among the public and private sector in 2016. *Malar J* 2017;**16**:494.

Muwanguzi J, Henriques G, Sawa P *et al.* Lack of K13 mutations in *Plasmodium falciparum* persisting after artemisinin combination therapy treatment of Kenyan children. *Malar J* 2016;**15**:36.

Mwai L, Kiara SM, Abdirahman A *et al.* In Vitro Activities of Piperaquine, Lumefantrine, and Dihydroartemisinin in Kenyan *Plasmodium falciparum* Isolates and Polymorphisms in *Pfcrt* and *Pfmdr1*. *Antimicrob Agents Chemother* 2009a;**53**:5069–73.

Mwai L, Ochong E, Abdirahman A *et al.* Chloroquine resistance before and after its withdrawal in Kenya. *Malar J* 2009b;**8**:106.

Mwangi TW, Ross A, Snow RW *et al.* Case Definitions of Clinical Malaria under Different Transmission Conditions in Kilifi District, Kenya. *J Infect Dis* 2005;**191**:1932–9.

Mwesigwa J, Achan J, Affara M *et al.* Mass Drug Administration With Dihydroartemisinin-piperaquine and Malaria Transmission Dynamics in The Gambia: A Prospective Cohort Study. *Clin Infect Dis* 2019;**69**:278–86.

Nabet C, Doumbo S, Jeddi F *et al.* Genetic diversity of *Plasmodium falciparum* in human malaria cases in Mali. *Malar J* 2016;**15**:353.

Nankabirwa J, Wandera B, Kiwanuka N *et al.* Asymptomatic Plasmodium infection and cognition among primary schoolchildren in a high malaria transmission setting in

Uganda. *Am J Trop Med Hyg* 2013;**88**:1102–8.

Ndungu FM, Marsh K, Fegan G *et al.* Identifying children with excess malaria episodes after adjusting for variation in exposure: identification from a longitudinal study using statistical count models. *BMC Med* 2015;**13**:183.

Neafsey DE, Juraska M, Bedford T *et al.* Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med* 2015;**373**:2025–37.

Njama-Meya D, Kamya MR, Dorsey G. Asymptomatic parasitaemia as a risk factor for symptomatic malaria in a cohort of Ugandan children. *Trop Med Int Heal* 2004;**9**:862–8.

Njuguna P, Maitland K, Nyaguara A *et al.* Observational study: 27 years of severe malaria surveillance in Kilifi, Kenya. *BMC Med* 2019;**17**:124.

Nkhoma SC, Banda RL, Khoswe S *et al.* Intra-host dynamics of co-infecting parasite genotypes in asymptomatic malaria patients. *Infect Genet Evol* 2018;**65**:414–24.

Nkhoma SC, Nair S, Al-Saai S *et al.* Population genetic correlates of declining transmission in a human pathogen. *Mol Ecol* 2013;**22**:273–85.

Noedl H, Se Y, Schaecher K *et al.* Evidence of artemisinin-resistant malaria in western Cambodia. *N Engl J Med* 2008;**359**:2619–20.

Nsobya SL, Parikh S, Kironde F *et al.* Molecular Evaluation of the Natural History of Asymptomatic Parasitemia in Ugandan Children. *J Infect Dis* 2004;**189**:2220–6.

O'Flaherty K, Maguire J, Simpson JA *et al.* Immunity as a predictor of anti-malarial treatment failure: a systematic review. *Malar J* 2017;**16**:158.

O'Meara WP, Bejon P, Mwangi TW *et al.* Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet* 2008;**372**:1555–62.

Obsomer V, Defourny P, Coosemans M. The *Anopheles dirus* complex: spatial distribution and environmental drivers. *Malar J* 2007;**6**:26.

Ofosu-Okyere A, Mackinnon MJ, Sowa MP *et al.* Novel Plasmodium falciparum clones and rising clone multiplicities are associated with the increase in malaria morbidity in

Ghanaian children during the transition into the high transmission season. *Parasitology* 2001;**123**:113–23.

Oguike MC, Falade CO, Shu E *et al.* Molecular determinants of sulfadoxine-pyrimethamine resistance in *Plasmodium falciparum* in Nigeria and the regional emergence of dhps 431V. *Int J Parasitol Drugs Drug Resist* 2016;**6**:220–9.

Okell LC, Bousema T, Griffin JT *et al.* Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. *Nat Commun* 2012;**3**:1237.

Okell LC, Ghani AC, Lyons E *et al.* Submicroscopic Infection in Plasmodium falciparum-Endemic Populations: A Systematic Review and Meta-Analysis. *J Infect Dis* 2009;**200**:1509–17.

Okell LC, Griffin JT, Kleinschmidt I *et al.* The Potential Contribution of Mass Treatment to the Control of Plasmodium falciparum Malaria. White NJ (ed.). *PLoS One* 2011;**6**:e20179.

Okell LC, Griffin JT, Roper C. Mapping sulphadoxine-pyrimethamine-resistant Plasmodium falciparum malaria in infected humans and in parasite populations in Africa. *Sci Rep* 2017;**7**:7389.

Okell LC, Reiter LM, Ebbe LS *et al.* Emerging implications of policies on malaria treatment: genetic changes in the *Pfmdr-1* gene affecting susceptibility to artemether–lumefantrine and artesunate–amodiaquine in Africa. *BMJ Glob Heal* 2018;**3**:e000999.

Okombo J, Kamau AW, Marsh K *et al.* Temporal trends in prevalence of Plasmodium falciparum drug resistance alleles over two decades of changing antimalarial policy in coastal Kenya. *Int J Parasitol Drugs Drug Resist* 2014;**4**:152–63.

Okombo J, Zongo I, Gadalla N *et al.* The Polymorphic Linker Domain of pfmdr1 Is Associated with Resistance-Conferring Mutations in Plasmodium falciparum Populations from East and West Africa. *Antimicrob Agents Chemother* 2013;**57**:4595–8.

Olliaro PL, Yuthavong Y. An Overview of Chemotherapeutic Targets for Antimalarial Drug Discovery. *Pharmacol Ther* 1999;**81**:91–110.

Olotu A, Fegan G, Wambua J *et al.* Estimating Individual Exposure to Malaria Using Local Prevalence of Malaria Infection in the Field. *PLoS One* 2012;**7**:e32929.

Omedo I, Mogeni P, Rockett K *et al.* Geographic-genetic analysis of *Plasmodium falciparum* parasite populations from surveys of primary school children in Western Kenya. *Wellcome Open Res* 2017;**2**:29.

Orjih AU, Ryerse JS, Fitch CD. Hemoglobin catabolism and the killing of intraerythrocytic *Plasmodium falciparum* by chloroquine. *Experientia* 1994;**50**:34–9.

Osier FHA, Weedall GD, Verra F *et al.* Allelic diversity and naturally acquired allele-specific antibody responses to Plasmodium falciparum apical membrane antigen 1 in Kenya. *Infect Immun* 2010;**78**:4625–33.

Ouédraogo A, Tiono AB, Diarra A *et al.* The effects of a pre-season treatment with effective antimalarials on subsequent malaria morbidity in under five-year-old children living in high and seasonal malaria transmission area of Burkina Faso. *Trop Med Int Heal* 2010;**15**:1315–21.

Owusu-Agyei S, Binka F, Koram K *et al.* Does radical cure of asymptomatic *Plasmodium falciparum* place adults in endemic areas at increased risk of recurrent symptomatic malaria? *Trop Med Int Heal* 2002;**7**:599–603.

Oyedeji SI, Awobode HO, Anumudu C *et al.* Genetic diversity of Plasmodium falciparum isolates from naturally infected children in north-central Nigeria using the merozoite surface protein-2 as molecular marker. *Asian Pac J Trop Med* 2013;**6**:589–94.

Pamba A, Richardson ND, Carter N *et al.* Clinical spectrum and severity of hemolytic anemia in glucose 6-phosphate dehydrogenase–deficient children receiving dapsone. *Blood* 2012;**120**:4123–33.

Pasupureddy R, Verma S, Pant A *et al.* Crucial residues in falcipains that mediate hemoglobin hydrolysis. *Exp Parasitol* 2019;**197**:43–50.

Peterson DS, Walliker D, Wellems TE. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in *falciparum* malaria. *Proc Natl Acad Sci* 1988;**85**:9114–8.

Peto TJ, Tripura R, Lee SJ *et al.* Association between Subclinical Malaria Infection and

Inflammatory Host Response in a Pre-Elimination Setting. Luty AJF (ed.). *PLoS One* 2016;**11**:e0158656.

Picot S, Olliaro P, de Monbrison F *et al.* A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in *falciparum* malaria. *Malar J* 2009;**8**:89.

Pinkevych M, Petravic J, Bereczky S *et al.* Understanding the Relationship Between *Plasmodium falciparum* Growth Rate and Multiplicity of Infection. *J Infect Dis* 2014;**211**:1121–7.

Pluess B, Tanser FC, Lengeler C *et al.* Indoor residual spraying for preventing malaria. *Cochrane Database Syst Rev* 2010, DOI: 10.1002/14651858.CD006657.pub2.

Poirot E, Skarbinski J, Sinclair D *et al.* Mass drug administration for malaria. *Cochrane Database Syst Rev* 2013;**12**:CD008846.

Polley and Conway. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* 2001;**158**:1505–12.

Ponsuwanna P, Kochakarn T, Bunditvorapoom D *et al.* Comparative genome-wide analysis and evolutionary history of haemoglobin-processing and haem detoxification enzymes in malarial parasites. *Malar J* 2016;**15**:51.

Le Port A, Cot M, Etard J-F *et al.* Relation between *Plasmodium falciparum* asymptomatic infection and malaria attacks in a cohort of Senegalese children. *Malar J* 2008;**7**:193.

Portugal S, Tran TM, Ongoiba A *et al.* Treatment of Chronic Asymptomatic Plasmodium falciparum Infection Does Not Increase the Risk of Clinical Malaria Upon Reinfection. *Clin Infect Dis* 2017;**64**:645–53.

Pradel G, Frevert U. Malaria sporozoites actively enter and pass through rat Kupffer cells prior to hepatocyte invasion. *Hepatology* 2001;**33**:1154–65.

Price RN, Uhlemann A-C, Brockman A *et al.* Mefloquine resistance in *Plasmodium falciparum* and increased *Pfmdr1* gene copy number. *Lancet* 2004;**364**:438–47.

Prothero RM. Migration and malaria risk. *Health Risk Soc* 2001;**3**:19–38.

Pryce J, Richardson M, Lengeler C. Insecticide-treated nets for preventing malaria. *Cochrane Database Syst Rev* 2018;**2018**, DOI: 10.1002/14651858.CD000363.pub3.

Putaporntip C, Kuamsab N, Kosuwin R *et al.* Natural selection of K13 mutants of *Plasmodium falciparum* in response to artemisinin combination therapies in Thailand. *Clin Microbiol Infect* 2016;**22**:285.e1-285.e8.

Ranson H, Lissenden N. Insecticide Resistance in African *Anopheles* Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. *Trends Parasitol* 2016;**32**:187–96.

Reed MB, Saliba KJ, Caruana SR *et al.* Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature* 2000;**403**:906–9.

Rice BL, Golden CD, Anjaranirina EJG *et al.* Genetic evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission. *Malar J* 2016;**15**:596.

Roca-Feltrer A, Carneiro I, Smith L *et al.* The age patterns of severe malaria syndromes in sub-Saharan Africa across a range of transmission intensities and seasonality settings. *Malar J* 2010;**9**:282.

Roestenberg M, McCall M, Hopman J *et al.* Protection against a Malaria Challenge by Sporozoite Inoculation. *N Engl J Med* 2009;**361**:468–77.

Rono J, Osier FHA, Olsson D *et al.* Breadth of Anti-Merozoite Antibody Responses Is Associated With the Genetic Diversity of Asymptomatic *Plasmodium falciparum* Infections and Protection Against Clinical Malaria. *Clin Infect Dis* 2013;**57**:1409–16.

de Roode JC, Pansini R, Cheesman SJ *et al.* Virulence and competitive ability in genetically diverse malaria infections. *Proc Natl Acad Sci* 2005;**102**:7624–8.

Roper C, Pearce R, Nair S *et al.* Intercontinental spread of pyrimethamine-resistant malaria. *Science (80-)* 2004;**305**:1124.

Roper C, Richardson W, Elhassan IM *et al.* Seasonal changes in the *Plasmodium falciparum*

population in individuals and their relationship to clinical malaria: a longitudinal study in a Sudanese village. *Parasitology* 1998;**116 (Pt 6**:501–10.

Ross LS, Dhingra SK, Mok S *et al.* Emerging Southeast Asian PfCRT mutations confer Plasmodium falciparum resistance to the first-line antimalarial piperaquine. *Nat Commun* 2018;**9**:3314.

Ross LS, Fidock DA. Elucidating Mechanisms of Drug-Resistant Plasmodium falciparum. *Cell Host Microbe* 2019;**26**:35–47.

Roucher C, Rogier C, Sokhna C *et al.* A 20-Year Longitudinal Study of Plasmodium ovale and *Plasmodium malariae* Prevalence and Morbidity in a West African Population. Braga ÉM (ed.). *PLoS One* 2014;**9**:e87169.

Royston P, Altman DG. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Appl Stat* 1994;**43**:429.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 2017, DOI: 10.1093/molbev/msx248.

RTS-S Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet* 2015;**386**:31–45.

Sagara I, Beavogui AH, Zongo I *et al.* Safety and efficacy of re-treatments with pyronaridine-artesunate in African patients with malaria: a substudy of the WANECAM randomised trial. *Lancet Infect Dis* 2016;**16**:189–98.

Sagara I, Beavogui AH, Zongo I *et al.* Pyronaridine–artesunate or dihydroartemisinin–piperaquine versus current first-line therapies for repeated treatment of uncomplicated malaria: a randomised, multicentre, open-label, longitudinal, controlled, phase 3b/4 trial. *Lancet* 2018;**391**:1378–90.

Salam N, Mustafa S, Hafiz A *et al.* Global prevalence and distribution of coinfection of malaria, dengue and chikungunya: a systematic review. *BMC Public Health* 2018;**18**:710.

Saliba KJ, Folb PI, Smith PJ. Role for the *Plasmodium falciparum* Digestive Vacuole in

Chloroquine Resistance. *Biochem Pharmacol* 1998;**56**:313–20.

Sanchez CP, McLean JE, Rohrbach P *et al.* Evidence for a *Pf*crt-associated chloroquine efflux system in the human malarial parasite *Plasmodium falciparum*. *Biochemistry* 2005;**44**:9862–70.

Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 1975;**94**:441–8.

Sauerbrei W, Royston P. Building Multivariable Prognostic and Diagnostic Models: Transformation of the Predictors by Using Fractional Polynomials. *J R Stat Soc Ser A (Statistics Soc* 1999;**162**:71–94.

Scopel KKG, Fontes CJF, Nunes ÁC *et al.* High prevalence of *Plamodium malariae* infections in a Brazilian Amazon endemic area (Apiacás-Mato Grosso State) as detected by polymerase chain reaction. *Acta Trop* 2004;**90**:61–4.

Sepúlveda N, Phelan J, Diez-Benavente E *et al.* Global analysis of *Plasmodium falciparum* histidine-rich protein-2 (*pfhrp2*) and *pfhrp3* gene deletions using whole-genome sequencing data and meta-analysis. *Infect Genet Evol* 2018;**62**:211–9.

Shretta R, Omumbo J, Rapuoda B *et al.* Using evidence to change antimalarial drug policy in Kenya. *Trop Med Int Heal* 2000;**5**:755–64.

Siddiqui FA, Cabrera M, Wang M *et al.* Plasmodium falciparum Falcipain-2a Polymorphisms in Southeast Asia and Their Association With Artemisinin Resistance. *J Infect Dis* 2018;**218**:434–42.

Sievers F, Higgins DG. Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. *Methods in Molecular Biology*. 2014, 105–16.

Sifft KC, Geus D, Mukampunga C *et al.* Asymptomatic only at first sight: malaria infection among schoolchildren in highland Rwanda. *Malar J* 2016;**15**:553.

De Silva PM, Marshall JM. Factors Contributing to Urban Malaria Transmission in Sub-Saharan Africa: A Systematic Review. *J Trop Med* 2012;**2012**:1–10.

Silvestrini F, Alano P, Williams JL. Commitment to the production of male and female gametocytes in the human malaria parasite *Plasmodium falciparum*. *Parasitology* 2000;**121**:465–71.

Sinden RE. Malaria, sexual development and transmission: retrospect and prospect. *Parasitology* 2009;**136**:1427–34.

Sinha A, Hughes KR, Modrzynska KK *et al.* A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* 2014;**507**:253–7.

Sinka ME, Bangs MJ, Manguin S *et al.* The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasit Vectors* 2010;**3**:117.

Sirawaraporn W, Sathitkul T, Sirawaraporn R *et al.* Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. *Proc Natl Acad Sci* 1997;**94**:1124–9.

Sisowath C, Petersen I, Veiga MI *et al.* In Vivo Selection of Plasmodium falciparum Parasites Carrying the Chloroquine-Susceptible pfcrt K76 Allele after Treatment with Artemether-Lumefantrine in Africa. *J Infect Dis* 2009;**199**:750–7.

Sisowath C, Strömberg J, Mårtensson A *et al.* In vivo selection of Plasmodium falciparum pfmdr1 86N coding alleles by artemether-lumefantrine (Coartem). *J Infect Dis* 2005;**191**:1014–7.

Sisya TJ, Kamn'gona RM, Vareta JA *et al.* Subtle changes in *Plasmodium falciparum* infection complexity following enhanced intervention in Malawi. *Acta Trop* 2015;**142**:108–14.

Smith TG, Lourenço P, Carter R *et al.* Commitment to sexual differentiation in the human malaria parasite, *Plasmodium falciparum*. *Parasitology* 2000;**121**:S0031182099006265.

Snow RW, Kibuchi E, Karuri SW *et al.* Changing Malaria Prevalence on the Kenyan Coast since 1974: Climate, Drugs and Vector Control. Keating JA (ed.). *PLoS One* 2015;**10**:e0128792.

Somé AF, Bazié T, Zongo I *et al. Plasmodium falciparum msp1* and *msp2* genetic diversity and allele frequencies in parasites isolated from symptomatic malaria patients in Bobo-Dioulasso, Burkina Faso. *Parasit Vectors* 2018;**11**:323.

Sondén K, Doumbo S, Hammar U *et al.* Asymptomatic Multiclonal Plasmodium falciparum Infections Carried Through the Dry Season Predict Protection Against Subsequent Clinical Malaria. *J Infect Dis* 2015;**212**:608–16.

Sondén K, Wyss K, Jovel I *et al.* High Rate of Treatment Failures in Nonimmune Travelers Treated With Artemether-Lumefantrine for Uncomplicated *Plasmodium falciparum* Malaria in Sweden: Retrospective Comparative Analysis of Effectiveness and Case Series. *Clin Infect Dis* 2017;**64**:199–206.

Sondo P, Derra K, Lefevre T *et al.* Genetically diverse *Plasmodium falciparum* infections, within-host competition and symptomatic malaria in humans. *Sci Rep* 2019;**9**:127.

Spanakos G, Snounou G, Pervanidou D *et al.* Genetic Spatiotemporal Anatomy of *Plasmodium vivax* Malaria Episodes in Greece, 2009–2013. *Emerg Infect Dis* 2018;**24**:541–8.

Stump AD, Atieli FK, Vulule JM *et al.* Dynamics of the pyrethroid knockdown resistance allele in western Kenyan populations of *Anopheles gambiae* in response to insecticide-treated bed net trials. *Am J Trop Med Hyg* 2004;**70**:591–6.

Sturm A, Amino R, van de Sand C *et al.* Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science (80-)* 2006;**313**:1287–90.

Sutherland CJ, Lansdell P, Sanders M *et al.* Pfk13-Independent Treatment Failure in Four Imported Cases of Plasmodium falciparum Malaria Treated with Artemether-Lumefantrine in the United Kingdom. *Antimicrob Agents Chemother* 2017;**61**, DOI: 10.1128/AAC.02382-16.

Takala-Harrison S, Jacob CG, Arze C *et al.* Independent Emergence of Artemisinin Resistance Mutations Among *Plasmodium falciparum* in Southeast Asia. *J Infect Dis* 2015;**211**:670–9.

Takala SL, Coulibaly D, Thera M a *et al.* Extreme Polymorphism in a Vaccine Antigen and Risk of Clinical Malaria: Implications for Vaccine Development. *Sci Transl Med* 2009;**1**:2ra5-2ra5.

Taylor SM, Parobek CM, Aragam N *et al.* Pooled deep sequencing of Plasmodium falciparum

isolates: an efficient and scalable tool to quantify prevailing malaria drug-resistance genotypes. *J Infect Dis* 2013;**208**:1998–2006.

Taylor SM, Parobek CM, DeConti DK *et al.* Absence of Putative Artemisinin Resistance Mutations Among *Plasmodium falciparum* in Sub-Saharan Africa: A Molecular Epidemiologic Study. *J Infect Dis* 2015;**211**:680–8.

Tetteh KKA, Stewart LB, Ochola LI *et al.* Prospective Identification of Malaria Parasite Genes under Balancing Selection. Doolan DL (ed.). *PLoS One* 2009;**4**:e5568.

Therneau TM. A Package for Survival Analysis in R. 2020.

Tiono AB, Guelbeogo MW, Sagnon NF *et al.* Dynamics of malaria transmission and susceptibility to clinical malaria episodes following treatment of Plasmodium falciparum asymptomatic carriers: results of a cluster-randomized study of community-wide screening and treatment, and a parallel entomology. *BMC Infect Dis* 2013;**13**:535.

Torre A della, Fanello C, Akogbeto M *et al.* Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. *Insect Mol Biol* 2001;**10**:9–18.

Torrentino-Madamet M, Fall B, Benoit N *et al.* Limited polymorphisms in k13 gene in Plasmodium falciparum isolates from Dakar, Senegal in 2012–2013. *Malar J* 2014;**13**:472.

Touré F, Bisseye C, Mavoungou E *et al.* Imbalanced distribution of *Plasmodium falciparum* EBA-175 genotypes related to clinical status in children from Bakoumba, Gabon. *Clin Med Res* 2006;**4**:7–11.

Trape JF, Pison G, Preziosi MP *et al.* Impact of chloroquine resistance on malaria mortality. *Comptes Rendus l'Academie des Sci - Ser III* 1998;**321**:689–97.

Trenholme GM. Treatment of patients with recrudescent falciparum malaria with a sequential combination of artesunate and mefloquine. *Infect Dis Clin Pract* 1993;**2**:220–1.

Triglia T, Menting JGT, Wilson C *et al.* Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in Plasmodium falciparum. *Proc Natl Acad Sci* 1997;**94**:13944–9.

Triglia T, Wang P, Sims PFG *et al.* Allelic exchange at the endogenous genomic locus in *Plasmodium falciparum* proves the role of dihydropteroate synthase in sulfadoxine-resistant malaria. *EMBO J* 1998;**17**:3807–15.

UNITAID. *Global Malaria Diagnostic and Artemisinin Treatment Commodities Demand Forecast (Phase 2)*., 2018.

de Valk HA, Meis JFGM, Bretagne S *et al.* Interlaboratory reproducibility of a microsatellite-based typing assay for *Aspergillus fumigatus* through the use of allelic ladders: proof of concept. *Clin Microbiol Infect* 2009;**15**:180–7.

de Valk HA, Meis JFGM, Klaassen CHW. Microsatellite based typing of *Aspergillus fumigatus*: Strengths, pitfalls and solutions. *J Microbiol Methods* 2007;**69**:268–72.

Veiga MI, Ferreira PE, Jörnhagen L *et al.* Novel polymorphisms in Plasmodium falciparum ABC transporter genes are associated with major ACT Antimalarial drug resistance. *PLoS One* 2011;**6**, DOI: 10.1371/journal.pone.0020212.

Veiga MI, Osório NS, Ferreira PE *et al.* Complex polymorphisms in the *Plasmodium falciparum* multidrug resistance protein 2 gene and its contribution to antimalarial response. *Antimicrob Agents Chemother* 2014;**58**:7390–7.

Verma AK, Bharti PK, Das A. HRP-2 deletion: a hole in the ship of malaria elimination. *Lancet Infect Dis* 2018;**18**:826–7.

Viriyakosol S, Siripoon N, Petcharapirat C *et al.* Genotyping of *Plasmodium falciparum* isolates by the polymerase chain reaction and potential uses in epidemiological studies. *Bull World Health Organ* 1995;**73**:85–95.

Vythilingam I, NoorAzian YM, Huat T *et al. Plasmodium knowlesi* in humans, macaques and mosquitoes in peninsular Malaysia. *Parasit Vectors* 2008;**1**:26.

Wamae K, Wambua J, Nyangweso G *et al.* Transmission and Age Impact the Risk of Developing Febrile Malaria in Children with Asymptomatic *Plasmodium falciparum* Parasitemia. *J Infect Dis* 2019;**219**:936–44.

Wang J, Zhang C-J, Chia WN *et al.* Haem-activated promiscuous targeting of artemisinin in

Plasmodium falciparum. *Nat Commun* 2015a;**6**:10111.

Wang Z, Wang Y, Cabrera M *et al.* Artemisinin Resistance at the China-Myanmar Border and Association with Mutations in the K13 Propeller Gene. *Antimicrob Agents Chemother* 2015b;**59**:6952–9.

Wargo AR, Huijben S, de Roode JC *et al.* Competitive release and facilitation of drug-resistant parasites after therapeutic chemotherapy in a rodent malaria model. *Proc Natl Acad Sci* 2007;**104**:19914–9.

Weiss DJ, Lucas TCD, Nguyen M *et al.* Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. *Lancet* 2019;**6736**:1–10.

White N. Antimalarial drug resistance and combination chemotherapy. Anderson RM (ed.). *Philos Trans R Soc London Ser B Biol Sci* 1999;**354**:739–49.

White NJ. Assessment of the pharmacodynamic properties of antimalarial drugs in vivo. *Antimicrob Agents Chemother* 1997;**41**:1413–22.

White NJ. Malaria parasite clearance. *Malar J* 2017;**16**:88.

White NJ, Van Vugt M, Ezzet F. Clinical pharmacokinetics and pharmacodynamics of artemether-lumefantrine. *Clin Pharmacokinet* 1999;**37**:105–25.

WHO. Severe falciparum malaria. World Health Organization. *Trans R Soc Trop Med Hyg* 2000;**94**:1–90.

WHO. *Methods for Surveillance of Antimalarial Drug Efficacy*. World Health Organization. https://www.who.int/malaria/publications/atoz/9789241597531/en/, 2009.

WHO. *Management of Severe Malaria. A Practical Handbook*. Third. World Health Organisation. https://www.who.int/malaria/publications/atoz/9789241548526/en/, 2013.

WHO. Severe Malaria. *Trop Med Int Heal* 2014a;**19 Suppl 1**:7–131.

WHO. *WHO Policy Brief for the Implementation of Intermittent Preventive Treatment of*

Malaria in Pregnancy Using Sulfadoxine-Pyrimethamine (IPTp-SP). World Health Organisation. https://www.who.int/malaria/publications/atoz/policy_brief_iptp_sp_policy_recommendation/en/, 2014b.

WHO. *Guidelines for the Treatment of Malaria*. 3rd ed., 2015.

WHO. *Test Procedures for Insecticide Resistance Monitoring in Malaria Vector Mosquitoes*. Second. World Health Organisation. https://www.who.int/malaria/publications/atoz/9789241511575/en/, 2016.

WHO. *A Framework for Malaria Elimination*. World Health Organisation. https://www.who.int/malaria/publications/atoz/9789241511988/en/, 2017.

WHO. *World Malaria Report 2018*. World Health Organization. https://www.who.int/malaria/publications/world-malaria-report-2018/report/en/, 2018a.

WHO. *Status Report on Artemisinin Resistance and ACT Efficacy*. World Health Organisation. https://www.who.int/malaria/publications/atoz/artemisinin-resistance-august2018/en/, 2018b.

WHO. *World Malaria Report 2019*. World Health Organization. https://www.who.int/malaria/publications/world-malaria-report-2019/en/, 2019.

WHO Global Malaria Programme. *The Role Of Mass Drug Administration, Mass Screening And Treatment and Focal Screening and Treatment For Malaria. Who Reference Number: Who/Htm/Gmp/2015.8*. World Health Organization., 2015.

Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016.

Wickham H, François R, Henry L *et al.* dplyr: A Grammar of Data Manipulation. 2018.

Wilson C, Serrano A, Wasley A *et al.* Amplification of a gene related to mammalian *mdr* genes in drug-resistant *Plasmodium falciparum*. *Science (80-)* 1989;**244**:1184–6.

Witkowski B, Amaratunga C, Khim N *et al.* Novel phenotypic assays for the detection of

artemisinin-resistant *Plasmodium falciparum* malaria in Cambodia: in-vitro and ex-vivo drug-response studies. *Lancet Infect Dis* 2013;**13**:1043–9.

Witkowski B, Duru V, Khim N *et al.* A surrogate marker of piperaquine-resistant *Plasmodium falciparum* malaria: a phenotype–genotype association study. *Lancet Infect Dis* 2017;**17**:174–83.

Wong W, Bai XC, Sleebs BE *et al.* Mefloquine targets the *Plasmodium falciparum* 80S ribosome to inhibit protein synthesis. *Nat Microbiol* 2017;**2**, DOI: 10.1038/nmicrobiol.2017.31.

Wongsrichanalai C, Barcus MJ, Muth S *et al.* A review of malaria diagnostic tools: Microscopy and rapid diagnostic test (RDT). *Am J Trop Med Hyg* 2007;**77**:119–27.

Woodland JG, Hunter R, Smith PJ *et al.* Chemical Proteomics and Super-resolution Imaging Reveal That Chloroquine Interacts with *Plasmodium falciparum* Multidrug Resistance-Associated Protein and Lipids. *ACS Chem Biol* 2018;**13**:2939–48.

Woodrow CJ, Dahlström S, Cooksey R *et al.* High-Throughput Analysis of Antimalarial Susceptibility Data by the WorldWide Antimalarial Resistance Network (WWARN) In Vitro Analysis and Reporting Tool. *Antimicrob Agents Chemother* 2013;**57**:3121–30.

Wootton JC, Feng X, Ferdig MT *et al.* Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature* 2002;**418**:320–3.

Wu L, van den Hoogen LL, Slater H *et al.* Comparison of diagnostics for the detection of asymptomatic *Plasmodium falciparum* infections to inform control and elimination strategies. *Nature* 2015;**528**:S86–93.

WWARN K13 Genotype-Phenotype Study Group. Association of mutations in the Plasmodium falciparum Kelch13 gene (Pf3D7_1343700) with parasite clearance rates after artemisinin-based treatments-a WWARN individual patient data meta-analysis. *BMC Med* 2019;**17**:1.

Xie SC, Dogovski C, Hanssen E *et al.* Haemoglobin degradation underpins the sensitivity of early ring stage Plasmodium falciparum to artemisinins. *J Cell Sci* 2016;**129**:406–16.

Yeka A, Kigozi R, Conrad MD *et al.* Artesunate/Amodiaquine Versus Artemether/Lumefantrine

for the Treatment of Uncomplicated Malaria in Uganda: A Randomized Trial. *J Infect Dis* 2016;**213**:1134–42.

Zhang Y, Meshnick SR. Inhibition of *Plasmodium falciparum* dihydropteroate synthetase and growth in vitro by sulfa drugs. *Antimicrob Agents Chemother* 1991;**35**:267–71.

Zhou M, Liu Q, Wongsrichanalai C *et al.* High prevalence of *Plasmodium malariae* and Plasmodium ovale in malaria patients along the Thai-Myanmar border, as revealed by acridine orange staining and PCR-based diagnoses. *Trop Med Int Heal* 1998;**3**:304–12.

Zhu SJ, Almagro-Garcia J, McVean G. Deconvolution of multiple infections in *Plasmodium falciparum* from high throughput sequencing data. Birol I (ed.). *Bioinformatics* 2018;**34**:9–15.

Zimmerman PA. *Plasmodium vivax* Infection in Duffy-Negative People in Africa. *Am J Trop Med Hyg* 2017;**97**:636–8.